Removing the Blinds on Injustice

How Privilege Interventions can Leverage Morality to Motivate Engagement in Social Change

Dissertation

kumulativ

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer.nat.)

des Fachbereichs Psychologie der Philipps-Universität Marburg

Vorgelegt von

Frank Eckerle, M.Sc.

Aus Mülheim an der Ruhr

Marburg, 2023

Vom Fachbereich Psychologie der Philipps-Universität Marburg
(Hochschulkennziffer 1180) als Dissertation angenommen am

Erstgutachter(in): Prof. Dr. J. Christopher Cohrs

Zweitgutachter(in): Dr. Maja Kutlaca

Tag der Disputation:

Acknowledgements

I am grateful for the extremely supportive research community in Marburg which allowed me to pursuit my curiosity with as little boundaries as possible. Over the years I made many bad and many good decisions in that sandbox, and so it was essential that I could always count on my colleagues, supervisors, and advisors to help me figure out which are which. I owe a lot to Christoher's ability to unravel complex questions without simplifying them, and for teaching me that there is fun to be had in admiring a problem together. I also benefitted a great amount of Uli's patience to continuously remind me that the first step is usually done before the second (although I feel like I am still not entirely convinced). And I am especially thankful for having met Maja on the way, who put a lot of her time and energy in my writing skills and has the unique ability to criticize work in a way that makes one believe in it more and not less. I also want to thank the PhD self-support group with Carmen, Inga, Teresa, and Adrian, who provided much needed confidence to start *doing* the research I did. The ideas presented in this thesis have also undergone a rigorous pretest procedure during many Rheder-hazed debates with Tijana, Eddie, Larissa, and Francesca, who are fountains of inspiration. On the office-side of things, there is of course Michaela and the many student assistants over the years who have supported me in so many ways with their work - or just by prolonging coffee breaks that led to that missing aha-effect. Thank you all!

An incredible source of energy and inspiration comes from all the time I spend with my dearest friends Nico and Lukas from my band Apewards. The once-a-week dive into our unreasonably loud cosmos kept me going through all else. I also want to thank Doreen for her unwavering effort to make me believe in the strengths within me.

The privilege to do this all comes from my parents, who built the fundament on which I can experience the kind of personal freedom I have.

Summary

In this cumulative dissertation I examine how people from advantaged groups (i.e., White people, men) react to situations where the unearned advantages they receive from their group membership are being pointed out to them. Such privilege confrontations or interventions have recently received a lot of attention in social psychological literature and the issue has become an increasingly hot topic in political discourse. One hope of researchers and practitioners in this field is to inspire support for pro-egalitarian policies among those who profit from inequality. However, pointing out privilege often triggers defensive reactions. In this dissertation I therefore tackled the important question: How can we point out that inequality is unjust to those who profit from it?

The starting point of this dissertation was the realization that privilege interventions have largely been studied as confrontations with evidence of personal privilege, which reflects an inappropriately individualized perspective on the systemic phenomenon that is inequality. Instead, I argue that research on privilege confrontations should examine how profiting from inequality can be communicated as a *moral* issue, thereby transforming systemic inequality into systemic *injustice*. Past research in the field has not explicitly investigated the role of morality, which is surprising because morality sits at the heart of pro-egalitarian motivation among political allies. The two manuscripts reported in this dissertation therefore examine two strategies how privilege confrontations might increase the moral relevance of unearned advantage.

Manuscript 1 was inspired by research on confrontations with sexism, which showed that the positionality of the confronter (male or female) affected men's responses to such confrontations. We therefore investigated whether men react differently to a male or female confronter of male privilege. Drawing from previous research on confrontations with ingroup transgressions, we examined specifically whether the confronter's gender affected the amount of image appraisal

(i.e., threat to the ingroup's image) and moral appraisal (i.e., threat to the ingroup's moral essence). Across three experiments with mixed-method design, we developed and tested a new appraisal-based model. We found that the female confronter triggered both more image *and* more moral appraisal, whereby image appraisal was associated with negative and moral appraisal was associated with positive attitudes toward male privilege.

Manuscript 2 was based on collective memory research, which indicates that confrontations with a negative collective past can inspire a moral growth mindset. As such, situating White privilege in national colonial history might trigger moral motives of social justice. We indeed found that this contextualized confrontation increased social justice motives, although the effect was confined to groups who are presumably more critically attached to Germany. Interestingly, none of the three experiments replicated the defensive effects of confronting with evidence of White privilege found in previous research from the US and UK. Specifically, confronting White Germans with White privilege did not affect how much life hardships they claimed. Taken together, Manuscript 2 indicates that White privilege can be communicated more effectively to some groups when it is situated in the origins of racial inequality and that reactions to White privilege confrontations are more strongly shaped by cultural differences, even within "Western" nations, than previous research had anticipated.

In total, the findings reported in this dissertation indicate the need to adapt some previous theoretical accounts on reactions to ingroup privilege. Specifically, they highlight the need and potential to investigate how moral considerations shape responses to privilege confrontations. There is also an urgent need to study in more detail how contextual factors shape the effects of confrontations with group-based privilege. More succinctly, the *who*, *how*, and *where* are important questions for future research to develop more effective privilege interventions that emphasize the systemic nature of social inequality to those who profit from it.

Zusammenfassung

In dieser kumulativen Dissertation untersuche ich, wie Menschen aus privilegierten Gruppen (d.h. Weiße, Männer) auf Situationen reagieren, in denen sie auf die unverdienten Vorteile hingewiesen werden, die sie durch ihre Gruppenzugehörigkeit genießen. Solche Konfrontationen mit Privilegien haben in der sozialpsychologischen Literatur zuletzt viel Aufmerksamkeit erregt, und das Thema ist auch im politischen Diskurs zunehmend in den Vordergrund gerückt. Eine Hoffnung von Forschenden und Praktiker*innen auf diesem Gebiet besteht darin, diejenigen, die von Ungleichheit profitieren, dazu zu bewegen, pro-egalitäre politische Maßnahmen zu unterstützen. Allerdings löst das Aufzeigen von Privilegien oft Abwehrreaktionen aus. In dieser Dissertation habe ich mich daher mit der wichtigen Frage beschäftigt: Wie können wir denjenigen, die von Ungleichheit profitieren, aufzeigen, dass diese ungerecht ist?

Ausgangspunkt dieser Dissertation war die Erkenntnis, dass Konfrontationen mit Privilegien weitgehend als Konfrontation mit Evidenz oder Beweisen für persönliche Privilegien untersucht wurden, was eine unangemessen individualisierte Perspektive auf das systemische Phänomen der Ungleichheit widerspiegelt. Stattdessen lege ich dar, warum Forschung zu solchen Interventionen stärker untersuchen sollte, wie das Profitieren von Ungleichheit als *moralisches* Problem kommuniziert werden kann, wodurch also systemische Ungleichheit in systemische *Ungerechtigkeit* transformiert werden kann. Die bisherige Forschung auf diesem Gebiet hat die Rolle von Moralität nicht explizit untersucht, was überraschend ist, da moralische Überlegungen zentral für die pro-egalitäre Motivation politischer Verbündeter („Allies“) sind. Die beiden Manuskripte in dieser Dissertation untersuchen daher zwei Strategien, wie Privilegienkonfrontationen die moralische Relevanz unverdienter Vorteile erhöhen können.

Manuskript 1 war inspiriert durch Forschung zu Konfrontationen mit Sexismus. Diese zeigte, dass die soziale Stellung der konfrontierenden Person (z.B. männlich oder weiblich) die

Reaktionen von Männern auf solche Konfrontationen beeinflusst. Wir untersuchten daher, ob Männer unterschiedlich auf Konfrontationen mit männlichen Privilegien reagieren, abhängig davon ob diese von männlichen oder weiblichen Personen kommen. In Anlehnung an Forschung zu Konfrontationen mit vergangenem Fehlverhalten der eigenen Gruppe, untersuchten wir, ob das Geschlecht der konfrontierenden Person das Ausmaß von Imagebewertung (d. h. Bedrohung durch negative Wahrnehmung der eigenen Gruppe durch andere) und moralischer Bewertung (d. h. Bedrohung der moralischen Essenz der eigenen Gruppe) beeinflusst. In drei Experimenten mit *mixed-method design* entwickelten und testeten wir ein neues, auf diesen unterschiedlichen Bewertungen basierendes Modell. Dabei fanden wir fanden heraus, dass die weibliche konfrontierende Person sowohl mehr Image- als auch mehr moralische Bewertungen von männlichen Privilegien auslöste, wobei Imagebewertung mit negativen Einstellungen und moralische Bewertung mit positiven Einstellungen gegenüber männlichen Privilegien zusammenhingen.

Manuskript 2 basierte auf Forschung zu *collective memory*, die darauf hinweist, dass Konfrontation mit einer negativen kollektiven Vergangenheit zu einer moralischen Wachstumsmentalität (*moral growth mindset*) führen kann. In diesem Sinne könnte die Kontextualisierung weißer Privilegien in der nationalen Kolonialgeschichte moralische Motive der sozialen Gerechtigkeit auslösen. Wir fanden tatsächlich, dass solch eine kontextualisierte Konfrontation die Motive für soziale Gerechtigkeit steigerte, wobei sich allerdings der Effekt auf Gruppen beschränkte, die vermutlich eine kritischere Beziehung zu ihrer Deutschen Nationalität haben. Interessanterweise replizierte keines der drei Experimente die Ergebnisse vergangener Forschung aus den USA und dem Vereinigten Königreich, welche eine defensive Reaktion auf Konfrontation mit Evidenz weißer Privilegien fand. Das heißt, weiße Deutsche scheinen im Vergleich weniger stark dazu zu neigen, vergangene Hindernisse und Schwierigkeiten im eigenen Leben hervorzuheben, um sich von weißen Privilegien zu distanzieren. Insgesamt deutet

Manuskript 2 darauf hin, dass weiße Privilegien einigen Gruppen gegenüber effektiver vermittelt werden können, wenn sie in den Ursprüngen rassischer Ungerechtigkeit verortet werden, und dass die Reaktionen auf Konfrontationen mit Privilegien stärker von kulturellen Unterschieden geprägt sind, als es die bisherige Forschung angenommen hat – sogar innerhalb "westlicher" Nationen.

Zusammengenommen deuten die in dieser Dissertation berichteten Ergebnisse darauf hin, dass einige frühere theoretische Darstellungen zu Reaktionen auf gruppenbasierte Privilegien angepasst werden müssen. Insbesondere unterstreichen die Ergebnisse die Notwendigkeit und das Potenzial, weiter zu verfolgen, wie moralischer Erwägungen die Reaktionen auf Privilegienkonfrontationen prägen. Außerdem erscheint es dringend erforderlich, genauer zu untersuchen, wie kontextbezogene Faktoren die Wirkung von Konfrontationen mit gruppenbezogenen Privilegien beeinflussen. Kurz gesagt, das *Wer*, *Wie* und *Wo* sind wichtige Fragen für künftige Forschung, die darauf abzielt, wirksamere Interventionen zu entwickeln, um die systemische Natur sozialer Ungleichheit bei denjenigen hervorzuheben, die von ihr profitieren.

Table of Contents

List of Figures

List of Tables

Introduction

Removing the "blinds" on injustice: How privilege interventions can leverage morality to motivate engagement in social change

Talking openly about power and privilege isn't easy, which is why people rarely do. The reason for this seems to be a fear of anything that might make dominant groups uncomfortable or "pits groups against each other" even though groups are already *pitted against one another by the structures of privilege that organize society as a whole. The fear keeps us from looking at what's going on and makes it impossible to do anything about the reality that lies deeper down.*

Allan G. Johnson (2006, p. 12, emphasis in the original)

Let me tell y'all what it's like,
Being male, middle-class and white.
It's a bitch, if you don't believe, listen up to my new CD.

Ben Folds – Rockin' the Suburbs

The distribution of power and resources in our globalized society is extremely unequal and resistant to change. Even though a growing outreach of political movements, such as the #MeToo and Black Lives Matter protests in recent years, has initiated a pro-egalitarian trend, inequality remains almost unchanged (Chancel et al., 2022). What contributes to this stark resilience of inequality?

One decisive factor lies in the capacity of systems of inequality (i.e., racism, sexism, classism etc.) to justify themselves. This is because groups and individuals that profit from inequality also tend to have the strongest influence in society (Helms, 2017). Therefore, advantaged groups can legitimize inequality through collective inaction: by simply not doing anything about it (Täuber & Moughalian, 2022). This also means that groups on the lower end of the social hierarchy are at a double disadvantage when it comes to fighting for social justice: First, they face a strong norm that supports the status quo and delegitimizes their claim for change (Burson & Godfrey, 2020; Jost & Hunyady, 2003). Second, even when they do manage to mobilize others to fight against

inequality, their effort might be in vain for as long as those in positions of power simply manage to ignore it.

The strength of this self-legitimizing power of inequality is enough to normalize even great atrocities of the past, which consequently are quickly forgotten and remain unrepaired. For example, after Haiti declared independence from France in 1804, the French government forced the new state, populated almost exclusively by former Black slaves, to pay a 150-million-franc indemnity to former French slave owners. This forced indemnity payment strangled Haiti's economy for over a century. Even though the sum (around 30 billion dollars in 2021) is small compared to France's economy, the French government never accepted Haiti's appeals to pause or pardon their payments (Piketty, 2022). The way in which the obvious injustice has gone ignored in this example is not an isolated case. A considerable amount of European (and generally "Western") wealth is built on such colonial legacy, rendering racism and White privilege an integral but invisible building block of European culture (Arndt, 2020; Bonds & Inwood, 2016). While political movements sometimes manage to achieve political concessions, such as the 2021 acknowledgement of the German genocide of the Herero and Nama between 1904-1908 (Deutscher Bundestag, 2023), fundamental changes are rare (the German government still rejects Namibia's call for legally binding reparations). This resilience to change is also not limited to racism, of course. For example, despite the numerous protests and movements to reduce the gender pay gap, globally, women still only make 35% of global work incomes. This is a meagre 4% increase since 1990 (Chancel et al., 2022). Inequality's resistance to change is so strong that some scholars have voiced "fundamental doubts about the modernizability of civil societies" (Waterstradt, 2018, p. 167).

What this brief overview tells us, is that a lack of critical consciousness prevents systemic inequality from being perceived as injustice and from being evaluated as immoral. *Critical consciousness* has originally been proposed as an important developmental factor that leads to a

delegitimization of status-supporting ideologies among minorities. The concept summarizes different psychological mechanisms that contribute to an appraisal of inequality as a systemic issue, which can motivate resistance and increase intraminority solidarity (Burson & Godfrey, 2020). The concept is less used for advantaged groups. However, research has identified in a similar vein that a reflection of unearned advantages can mobilize the privileged to act in solidarity with the disadvantaged (Bergkamp, Olson, & Martin, 2022; Goodman, 2011; Radke et al., 2020). Specifically, a critical reflection on the inequality in moral terms seems to be a crucial motivator behind political action against the status quo (Curtin et al., 2015; Thomas et al., 2022; Thomas et al., 2009).

Such political solidarity of the advantaged with the disadvantaged, in turn, is an important contributor to the success of social change efforts (Subašić et al., 2008). To draw from another example, there was a collective mobilization against the deportation of Bulgarian Jews in Bulgaria during World War II. An extensive analysis of contemporary documents indicated that the movement communicated resistance to the deportation of Bulgarian Jews as a moral stake, which may have contributed to the at least partially successful resistance against their deportation. This morality impetus is present in many of the analyzed speeches, open letters, and other documents. To quote one: “When the Bulgarian people lose their sense of justice that they have nurtured over the course of centuries and that is so much a part of their national identity, they will lose their moral and spiritual uniqueness, their Slavic essence, their Bulgarian face” (Reicher et al., 2006, p. 63). As can be seen from this example, the moral stakes did not extend to *non*-Bulgarian Jews from Bulgaria’s annexed territories in Greece and former Yugoslavia, who were still being deported and killed (Ragaru, 2017b). So while the story of the rescue of Bulgarian Jews is therefore much more complicated and less heroic than often portrayed (Ragaru, 2017a), it still illustrates how raising the stakes of morality can mobilize resistance against injustice (Skitka, 2010). This thesis will therefore focus on the role of morality in

confrontations with group-based privileges and how moralizing inequality is related to political solidarity.

What is group privilege?

All people belong to multiple groups and together these identities contribute to one's individual experience of privilege and discrimination, or in other words, one's relation to society (Case et al., 2012). Race- and gender-based inequality make up a large degree of this experience, because they are deeply related to systemic forces of inequality and social dominance (Pratto et al., 2006; Sidanius & Veniegas, 2000). Overall, the combination of groups a person belongs to contributes to a person's social position (Hearn, 2012). As such, the term "group privilege" is a way to describe the collection of advantages that come with privileged identities, such as being White, male, middle-class and so on (McIntosh, 2012).

People identify with their social categories and the extent of their identification can explain their inclination to act *as* members of these categories. More specifically, Social Identity Theory (Abrams & Hogg, 1990; Tajfel & Turner, 1979) posits that in situations where certain social categories become salient (e.g., a woman walks into a room filled with only men), people are motivated to protect and foster their personal self via their social identity (e.g., their gender). However, previous research indicates that threatening an ingroup by highlighting group privileges motivates members of that group to disengage from their social identity and to adopt an individualized reasoning of inequality. For example, men might react to confrontations with their male privilege by claiming that they had never personally profited from male privilege because their social status was a result of hard work. In the same prospect, White individuals might distance from White privilege by claiming that they had endured many personal hardships (L. T. Phillips & Lowery, 2015). Such individualization of inequality is a defensive strategy that supports the normative (i.e., socially accepted and unquestioned) avoidance of privilege (L. T. Phillips & Lowery, 2017).

It is true, however, that the presence of group privilege (in some domains) should not be equated with the absence of disadvantage (in other domains). It is more fruitful to think of privilege and disadvantage as originating from the same systems of inequality, such as racism, sexism, or classism (Nixon, 2019). Case et al. (2012, p. 4) summarized the relationship between privilege and oppression to be "inseparable as codependent structural forces", while Johnson (2006, p. 38) put it more vividly: "Just as privilege tends to open doors of opportunity, oppression tends to slam them shut". Privilege and oppression have also frequently been described as opposite sides of the same coin (Bergkamp, Martin, & Olson, 2022; Goodman, 2015; Nixon, 2019). While social psychological research has a long tradition of investigating inequality in the form of prejudice and discrimination of disadvantaged groups, research on the social psychology of privilege as well as on the psychological effects of living within systems of inequality has only recently begun to attract greater attention (Dixon, 2012; Leach & Teixeira, 2022).

Why should we look at group privilege?

There are several practical and epistemic benefits of investigating group privilege and how it relates to political action. First, the existence of group privilege is normalized and thus profiting from inequality is an invisible and automatic process (Barajas, 2022). This normalization of inequality is reinforced through the absence of "alternative visions for society" (King, 1991, p. 135) and through an understanding of privilege as independent from disadvantage, rather than as its flipside (Pratto & Stewart, 2012). As such, avoiding group privilege contributes to the justification of inequality, exemplified through terms such as *spiral of silence* (Noelle-Neumann, 1974), *herd invisibility* (L. T. Phillips & Lowery, 2017) or *cultivated ignorance* (Bergkamp, Olson, & Martin, 2022). Effectively pointing out group privilege therefore has the potential to break through the normative self-maintenance of inequality.

Second, because those who profit from systems of inequality have a lot of political influence, raising awareness of privilege is a needed pathway to motivate a stronger moralization of

inequality as unjust. Realizing that one's social position is related to systemic inequality is an important first step in counteracting status-maintaining ideologies such as meritocratic beliefs and in fostering a social justice ideal of "participational parity" (Thrift & Sugarman, 2019). This is illustrated nicely by a series of experiments, where inequality between the gardens of villagers of an imagined village was either emphasized to be a result of luck (i.e., random events that impair harvest) or of the individual effort the villager put into it. The experiments demonstrated that people, independent of their political orientation and ideology, prefer redistribution more when they perceive inequality to be rooted in luck (Nettle & Saxe, 2020). In a similar fashion, understanding one's advantaged starting position in systemic inequality to be determined by a birth lottery may lead to pro-egalitarian attitudes.

Third, and relatedly, creating awareness of group privileges can help taking some burden of "moral labor" from the disadvantaged, especially if the confrontation with privilege stems from another person of the privileged ingroup: As outlined above, disadvantaged groups that want to change the status quo have few options besides collectivizing and moralizing systemic inequality to convince an inert majority to support their cause for social change (Subašić et al., 2008). However, these repeated confrontations with the inequality and the majorities' continued resistance and inaction can lead to trauma and burn-out (Chen & Gorski, 2015; Gorski, 2019). In contrast, a reflection on privilege and the participation in collective action for social change is often perceived as a transformative and essentially positive experience among privileged allies (Goodman, 2011). But again, only when the advantaged become aware of their positionality can they provide efficient support in social change efforts (Radke et al., 2020; Spanierman & Smith, 2017). However, as we will see, it is not a trivial task to trigger such reflection.

Constructive and defensive responses to privilege confrontations

Most research on privilege confrontations has been conducted on male and White privilege and the current thesis also focuses on these two domains of group privilege. I will therefore

summarize the state of the literature on privilege confrontations with an emphasis on male and White privilege. However, at points where important additional insights come from research on other privilege domains, I will report them as well.

Most social psychological investigations of male and White privilege confrontations find that being confronted with these privileges has negative effects on self-esteem and leads to negative emotions. Some works find that this *can* explain constructive responses to privilege interventions. Both men and White individuals reported feelings of guilt after being confronted with their respective group privilege (Schmitt et al., 2000) and guilt and anger are associated with support for pro-egalitarian policies (Iyer et al., 2003; Leach et al., 2006). Men also reported lower levels of personal and collective self-esteem as well as less emotional attachment to their gender after being invited to think about their male privilege compared to male disadvantage (Branscombe, 1998). A similar reflective exercised was used to confront White people with their White privilege, leading to stronger feelings of collective guilt than thinking about Black disadvantage (Powell et al., 2005). In attempt to motivate a reflection on ingroup privilege, men (and heterosexuals) were presented a video on their group's privileges, which led them to report more privilege awareness and to reject modern sexism claims (claims that gender equality had already been achieved, like "Society has reached the point where women and men have equal opportunities for achievement"; Swim et al., 1995).

However, fewer positive effects were reported after asking White individuals to reflect on their White privilege (versus outgroup disadvantage or a race-irrelevant topic), which led to more modern racism among participants who highly identified with being White (Branscombe et al., 2007). White individuals also claimed more life hardships after being confronted with White privilege (compared to empty control conditions) and life hardships were in turn negatively related to privilege acceptance and support for pro-egalitarian policies (Murdoch & McAloney-Kocaman, 2019; L. T. Phillips & Lowery, 2015). However, after completing a self-affirmation

exercise, the relationship reversed (L. T. Phillips & Lowery, 2015). Life hardships have been interpreted as a motivated response to distance the self from ingroup privilege, because inequality presented as White privilege (compared to Black disadvantage) leads to a negative self-image (Lowery et al., 2007). Similarly, confrontations with class privilege led to claims of life hardships among middle and upper class individuals and increased attempts to defend personal status (L. T. Phillips & Lowery, 2020). Confrontations with evidence of White privilege also reduced collective self-esteem, which is thought to motivate action to repair the ingroups social image (Lowery et al., 2012). Self-image concerns were however negatively associated with White people's willingness to think of racism in systemic versus individual terms (Unzueta & Lowery, 2008).

An attempt to compare reactions of men and White people to confrontations with their privilege provided disillusioning evidence: reactions to confrontations seem to largely depend on participant's prior attitudes toward inequality. Specifically, those who tended to deny the existence of group-based inequality also responded more defensively than those who acknowledged inequality beforehand (Puryear et al., 2019). Other recent investigations on privilege intervention effectiveness have corroborated these results: Younger and more left-wing participants accepted White privileges more after the confrontation whereas older and more conservative participants accepted them less (Brad et al., 2019). Unsurprisingly, White people with a strong system-justifying ideology were more likely to react defensively to confrontations with White privilege as well (Strupp-Levitsky, 2020). An important challenge for privilege confrontations thus seems to be to actually *change* attitudes toward inequality.

Promising insights come from White privilege confrontations that highlight the system of inequality that perpetuates privilege: For example, a diversity training focused on intersectionality has been shown to raise privilege awareness across multiple domains (Ehrke et al., 2020), a confrontation with experiences of Black discrimination increased privilege

awareness and political action (Uluğ & Tropp, 2021). Similarly, providing a systemic account of White privilege led to less colorblind attitudes, less modern racism as well as more affirmative action intentions than a framing of White privilege as the result of a human tendency for ingroup bias (Tittler et al., 2023). More generally, framing efforts to overcome racial injustice as a moral opportunity rather than a moral obligation led to more positive attitudes (Does et al., 2011) and more approach-based physiological responses (Does et al., 2012) in White individuals.

As can be seen from this brief literature review there are three rough patterns of results: First, confrontations trigger negative affect and threaten personal and collective self-esteem. Second, privilege confrontations often only exacerbate previously held attitudes toward inequality and work best among those who are already converted. Third, situating ingroup privilege firmly within systems of inequality seems to produce more consistent and positive effects. The 3D model provided a first attempt to explain the different mechanisms behind negative and positive effects of privilege confrontations (Knowles et al., 2014).

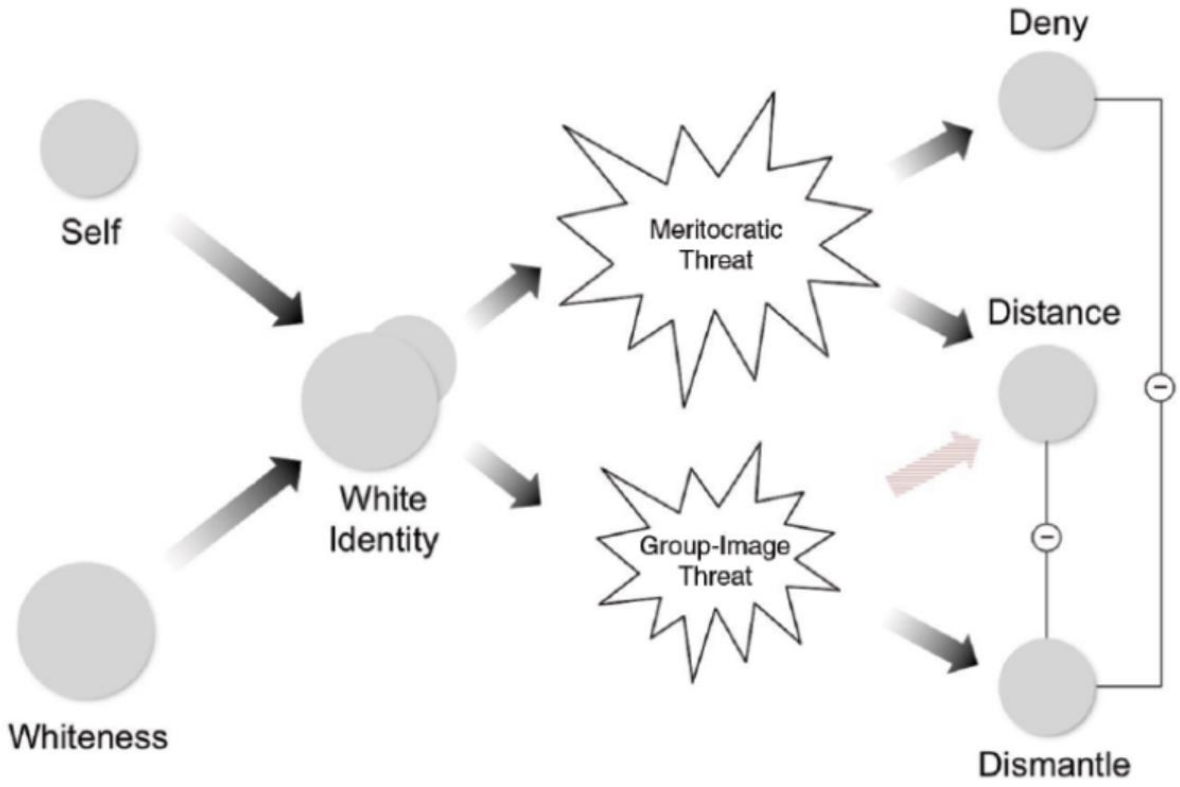
Explaining defensive and constructive responses: the 3D model and beyond

The 3D model (Knowles et al., 2014) argues that ingroup privilege is not invisible, but that it is actively avoided to counter identity threat. The model uses the example of White privilege to theorize that a salient privileged identity triggers two distinct forms of threat: threat to the ingroup's social image and to meritocratic ideology. These threats then cause three different reactions (the three "D's"): *Denying* the existence of privilege, *distancing* the self from the group or from ingroup privilege (i.e., "I have never profited from being White") or *dismantling* the underlying system of inequality.[1] Group-image threat is expected in the model to motivate dismantling behavior, because it "affects the collective self and has moral connotations".

[1] Arguably, the two defensive outcomes (deny and distance) are rooted in attempts to downplay the systemic nature of inequality. As such, the model also implicitly provides insight on the motivational basis behind system justification of the advantaged (Hässler et al., 2019). I will come back to this point in the conclusion.

Meritocratic threat instead "implicates the personal self and its competence" and should therefore motivate denial and distancing (Knowles et al., 2014, p. 598). The model assumes that meritocratic threat takes precedence over group-image threat, because saving one's personal sense of agency would be prioritized over managing the group's image. While focusing on research on White privilege in the US, the authors suggest that the 3D model is "[…] a cultural-psychological theory of responses to ingroup dominance in contexts that are Western, democratic, and individualistic." (Knowles et al., 2014, p. 604). In summary, the model predicts a positive effect through group-image threat, a negative effect through meritocratic threat, and a more primary role of meritocratic threat in Western contexts. The model is depicted in Figure 1. In the following I will outline some epistemological problems and empirical shortcomings of this model.

Figure 1. Visual representation of the 3D model



Note. This graphic is unchanged from the original article by Knowles et al. (2014).

Empirical and epistemological concerns regarding the 3D model

One main problem of the available data on reactions to privilege confrontations concerns their potential generalizability. The samples across the White and male privilege confrontation experiments that I summarized above (N = 4308) consist to 61% of US-Americans and only four of the publications used non-American samples, two of which sampled in Australia and the UK. This state of the literature speaks volumes to the need to investigate the psychology of privilege across more contexts and with less specific demographics. Because the 3D model draws to a large degree from evidence found in these samples but claims to be "a cultural-psychological theory of responses to ingroup dominance in contexts that are Western, democratic, and individualistic" (Knowles et al., 2014, p. 604), it should be approached with the appropriate caution. Still, the 3D model is highly influential in the field and has informed a lot of recent research on White privilege and beyond.

The model's main claims are essentially empirical hypotheses, but they have not been tested directly. The closest tests of the model I could find come from two publications on White privilege, one in the US (L. T. Phillips & Lowery, 2015) and one replication in the UK (Murdoch & McAloney-Kocaman, 2019). There is also a series of experiments on class privilege from the US that use a similar design (L. T. Phillips & Lowery, 2020). Overall, these studies seem to support the negative effect of meritocratic threat: They found that a confrontation with evidence of privilege triggered claims of life hardships and that hardship-claiming was associated with lower acceptance of privilege and less support for equality-enhancing policies. Completing a self-affirmation exercise to protect against threats to self-esteem (Sherman & Cohen, 2002; Steele, 1988) prior to the confrontation reduced participant's tendency to claim hardships and increased support for equality-enhancing policies in the confronted group. The findings are interpreted as evidence that privilege confrontations threaten positive self-esteem in the

advantaged and that claiming life hardships serves as a meritocratic justification that helps to create distance between the self and group-based privilege.

However, the life hardship studies have worrying empirical inconsistencies and limitations. In the US study on White privilege, for example, only one of the three experiments actually finds a significant effect of the confrontation with evidence of White privilege on life hardships (L. T. Phillips & Lowery, 2015): In one experiment, the effect was described to be significant even though the result was reported at $p = .057$ (Experiment 1b), and in another the confidence interval included zero (Experiment 2) and only the interaction between privilege confrontation and self-affirmation was significant (Experiment 2). In that experiment, all confidence intervals of indirect effects through life hardships on affirmative action also included zero (Experiment 2). The indirect effect was only significant when belief in White privilege was included as a second mediator after life hardships, even though belief in White privilege was originally hypothesized not to be affected by the confrontation. Additionally, in the UK study on White privilege, the effect of the White privilege confrontation on life hardships was only present when controlling for gender and still only explained 5% of variance in life hardships (Murdoch & McAloney-Kocaman, 2019). Results of a stepwise regression (the authors do not explain why they did not use a more direct approach to gauge whether the correlational pattern was indicative of mediation), were interpreted in a reversed order from the one in the US study: confrontation -> belief in White privilege -> life hardships. However, as soon as belief in *personal* White privilege was introduced to the model, only this variable remained significantly associated with life hardships. Finally, in the study on class privileges, the effect on life hardships was similarly inconsistent and qualified by several boundary conditions. For example, the effect was only significant as an indirect effect where self-esteem was the mediator. Life hardships were also not significantly related to high or low personal merit mindsets and not significantly affected by priming a hard versus easy life before the confrontation (L. T. Phillips & Lowery, 2020). All in

all, the lack of consistent effects challenges the 3D model's assumption that meritocratic threat is primary response to confrontations with ingroup privilege.

Still, evidence for group-image threat (the positive mediator proposed in the 3D model) is even thinner: According to the original article on the 3D model, one study found that White individuals who were confronted with White privilege favor "privilege-reducing" (Knowles et al., 2014, p. 603) strategies. However, a closer look reveals that confrontations with White advantage only led to support for policies that reduce economic opportunities of White people but *not* to support for policies that increase opportunities of minorities. Additionally, the White privilege confrontation did not lead to a specific motivation to "repairing the reputation of the racial ingroup" (Knowles et al., 2014, p. 603) but was associated with the broader concept of negative collective self-esteem (Lowery et al., 2012). At best these results are very indirect evidence for the model's contention that White people would dismantle systems of privilege *specifically to repair their group-image*. In fact, there is counter evidence which shows that reminders of ingroup transgressions are negatively associated with outgroup orientations when the transgression is perceived as a threat to the ingroup's image (Allpress et al., 2014).

The shortcomings of attempts to reduce defensiveness

Another problem of previous research on (White) privilege concerns the use of self-affirmation to test whether threat to self-esteem motivates defensiveness. In some experiments, self-affirmed White people were more open to receiving negative information about the ingroup, which led them to claim less life hardships, and to support affirmative action (L. T. Phillips & Lowery, 2015) as well as policies to reduce unfair advantages (Lowery et al., 2012). However, support for policies to support the opportunity of Black people were not affected and there is no research on the long-term effectiveness of pairing evidence confrontations with self-affirmation. It is thus unclear whether self-affirmation prior to a confrontation has positive effects beyond inducing a situational mood change that allows oneself to take over a "benevolent" stance on one's own

advantaged position. In this sense, the positive influence of self-affirmation might simply be a manifestation of the paradox of perceiving privilege: "[…] if you don't feel good, then you must not be privileged" (Johnson, 2006, p. 37). Self-affirmation can also backfire. For example affirmation of the moral domain can license immoral behavior (Sherman & Cohen, 2006). Adopting self-affirmation exercises in practice is therefore a dangerous road when we consider the importance of moral motives behind political solidarity and allyship (Kutlaca, Radke, et al., 2020; Radke et al., 2020).

These shortcomings illustrate how the focus on self-esteem risks falling into the hedonism trap of psychology (N. L. Phillips et al., 2015): (Social) psychological research often focusses much more on fostering the adjustment to negative circumstances than examining how negative emotional states can motivate transformative action (Allen & Leach, 2018; Thrift & Sugarman, 2019). It is important to remember that fostering the successful dissociation from injustice is exactly the opposite of what is needed when attempting to transform silence into action (Lorde, 2012).

Finally, the epistemological focus on managing self-esteem also reflects an individualization bias in social psychology, which often fails to account for how systemic processes influence affect (Leach & Bou Zeineddine, 2021) and how contextual, macro, or system-level factors contribute to the motivated avoidance of ingroup privilege (Leach & Teixeira, 2022; Täuber & Moughalian, 2022). By individualizing systemic mechanisms (i.e., as issues of self-esteem), social psychologists risk contributing to the maintenance of the status quo (N. L. Phillips et al., 2015; Teo, 2015). Instead, as Bergkamp, Martin, and Olson (2022, p. 24) argue, when group privilege is understood as "[…] an essential, implicit component of an oppressive system, […] awareness of social privilege can elucidate the mechanisms that scaffold inequity and ultimately serve a social justice mission".

Based on these criticisms, I argue that awareness of group privilege *should* feel uncomfortable and that such negative states can motivate social change (Allen & Leach, 2018; Leach, 2019). A similar argument has been made in research on collective memory, which seems to indicate that negative examples of the ingroup's history can foster a moral growth mindset, motivating action to redeem the ingroup's past (Lienen, 2023).

A more systemic research focus however should not allow individuals to disengage from injustice, because it is communicated as not their personal fault. Instead, such a research agenda should find ways to make individuals feel responsibility about injustice *even though* it is not their fault. For this we need more research that examines how contextual circumstances shape responses to confrontations with ingroup privilege: Who is the confronter? How is the confrontation embedded in systemic inequality? Which culturally entrenched ideologies are available that justify inequality? The projects reported in this thesis provide a first foray into these questions. Specifically, we investigated the effects of the confronter's positionality (male versus female) on men's reactions to their confrontation and of contextualizing White privilege in negative collective memory, namely national involvement in colonialism. This leads us to a central question: *What works* in privilege interventions? As introduced above, in this dissertation I focus on *morality* as an underexplored and potentially very constructive mechanism of privilege interventions: How can moral concerns over inequality be effectively triggered? How do such moral concerns relate to the acceptance of group-based privileges and to the motivation to dismantle systemic injustice?

The role of morality in privilege interventions

Morality is a primary dimension for the evaluation of ingroups (Ellemers et al., 2013; Leach et al., 2007). Morality regulates social behavior, because the blemish of immorality is highly threatening and can lead to rejection from society (Fitouchi et al., 2021; Gausel et al., 2012). This means that threatening information about the ingroup's morality usually produces strong

reactions, for example by derogating the source of that information (Kutlaca, Becker, & Radke, 2020; Minson & Monin, 2012) or by morally engaging with the threatening information and striving for redemption (Wenzel et al., 2020). As such, morality is a strong motivator for both backlash and for more constructive responses. A central switch for this effect is the presence of moral opportunity (van der Toorn et al., 2015): When people appraise instances of moral failure as a *specific self-defect* they are more likely to accept these failures and respond constructively whereas more general and non-repairable appraisals predict defensive reactions (Cibich et al., 2016; Gausel & Leach, 2011; Gausel et al., 2012; Vitriol & Moskowitz, 2021). Moreover, messages that threaten morality seem to be of greater concern when they come from fellow ingroup members (Kutlaca, Becker, & Radke, 2020).

Moral threat should be clearly distinguished from (group-)image threat. For example, reminders of ingroup transgressions can trigger two different dimensions of shame: one rooted in a concern to fail moral standards of the ingroup, the other rooted in a concern for the group's image. While moral shame was associated with positive outgroup orientations, image shame had negative effects (Allpress et al., 2014). Recent research revealed accordingly that the opposition of advantaged groups to collective action of low-status groups is rooted in image threat (Teixeira et al., 2020). This is also in line with the finding that concerns about negative ingroup evaluation can lead to *negative* intergroup outcomes, while concerns about failing to meet the moral standards of the ingroup are likely to have *positive* intergroup outcomes (Ellemers & van den Bos, 2012), especially when there is opportunity for redemption (Does et al., 2011; Does et al., 2012).

Taken together, privilege confrontations that threaten the morality of privilege but provide opportunity for moral redemption may have a high chance of being effective. Even if, or more precisely, *because* they will not feel good. The manuscripts reported below therefore approach the question how to effectively communicate to the privileged that systemic inequality is

immoral. The first manuscript examines the role of the positionality of the messenger and the second manuscript investigates the effect of drawing connections between past ingroup transgressions and present ingroup privilege.

Positionality of the confronter and moral versus image appraisal

Whereas the role of confronter positionality has received considerable attention in the literature on confrontations with prejudice (Czopp & Monteith, 2003; Gulker et al., 2013) there is almost no research on how the membership of the confronter in either the privileged or disadvantaged group affects reactions toward privilege confrontations. The position of the confronter is important, because when members of the privileged group start reflecting their privilege, they take a good amount of moral labor from the disadvantaged group. As outlined above, members of disadvantaged groups would otherwise need to confront the privileged with the inequality of the status quo themselves. However, it would be fruitful to know whether ingroup confronters are actually effective communicating privilege, especially considering the stark resistance to gender equality (Flood et al., 2021). I could find only one study where the effect of ingroup/outgroup positionality of a privilege confronter was manipulated. Results indicated that a White confronter induced greater acknowledgment of racial inequity than a Black confronter. However, both confronters were professors, which added an element of power between participants and the confronters that might affect the results (Kahalon et al., 2019).

Psychologically speaking, the positionality of the confronter may play an important role for whether their message is appraised as moral or image threat and moral and image appraisal seem to have different implications for subsequent reactions (Allpress et al., 2014). Advantaged groups often expect moral confirmation from disadvantaged outgroups (Kahalon et al., 2019; Siem et al., 2013). As such, privilege confronters from disadvantaged outgroups might be likely to trigger concerns regarding the ingroup's image. This is also evident in the backlash effect observed against female confronters of sexist behavior in men (Pease, 2021). Confronters from

within the privileged ingroup seem to be in a better position to highlight immorality, because they receive more trust and less suspicion about having ulterior motives (Drury & Kaiser, 2014). However, as outlined above this moral threat can also backfire against the confronter, leading to a phenomenon called Black Sheep Effect, where the ingroup messenger is derogated to diffuse their confrontation (Kutlaca, Becker, & Radke, 2020; Marques & Paez, 1994; Monin et al., 2008). Manuscript 1 therefore attempted to disentangle the relation between ingroup/outgroup positionality of the confronter and image/moral appraisal in the realm of male privilege.

Contextualizing privilege and it's potential to activate social justice morality and solidarity

There are instances where the systems of inequality that contribute to one's personal set of advantages become undeniable, especially in situations where one clearly profits from them or someone else clearly suffers from their absence. Some excerpts of an unpublished interview study with German advantaged group activists illustrate how situations that shine a new light on the status quo can be important moments of politization (Eckerle & Lienen, manuscript in preparation; see also van Zomeren et al., 2018).

> „Ich habe mit 16 mich angefreundet mit Asylbewerbern in meiner Stadt und war jeden Tag in deren Wohnhaus und habe dann ziemlich schnell gemerkt, dass es einfacher ist für die Menschen dort etwas auch bei Ämtern und Behörden durchzukriegen, wenn ich dabei bin. [When I was 16, I made friends with asylum seekers in my town and went to their homes every day. I realized pretty quickly that it's easier for the people there to get things through to the authorities if I'm there.]"
>
> „Ich glaube, dass es irgendwie so Informationen darüber braucht, dass andere Menschen in irgendeiner Form sagen: „Ja, das ist ja nicht so selbstverständlich, weil unsere Sicht ist eine ganz andere. [I think that it somehow needs information about this in such a way that other people say in some form: "Yes, that's not so self-evident, because our view is completely different.]"

These examples demonstrate the crucial role of wearing down the armor of normative avoidance of injustice through disruption, dissent and innovation (Moscovici & Mugny, 1983; Ozden, 2022; Shuman et al., 2021). For this it seems promising to focus on involuntary and unconscious practices through which inequality, meaning disadvantage *and* privilege, are maintained (Lucal,

1996). For example, male privilege could be communicated not simply as a throne that men allegedly sit on, but as a patriarchal system that efficiently divides male and female workers for the benefit of whoever profits from their work.[2] In another example, the realization that White supremacy and the ideology of racism was an invention that justified the invasion, exploitation, and destruction of native populations around the globe (Smedley & Smedley, 2005), might motivate White individuals to accept White privilege, moralize social inequality and consider allying with Black movements to abolish the remnants of this imperial project (Kelliher, 2018). Manuscript 2 therefore investigated the effects of contextualizing White privilege within the historically entrenched system of racial inequality.

Projects reported in this thesis

This dissertation aims to provide new vantage points to privilege interventions. Instead of managing defensive reactions, I propose to focus more strongly on the role of contextual factors (i.e., the source and framing of the confrontation) to derive a more thorough understanding of the psychological process behind defense against and acceptance of privilege, and to help design more efficient interventions (see also Leach & Teixeira, 2022). Ultimately, the two projects may contribute to our understanding of *what works in privilege interventions*. Specifically, how privilege confrontations can increase the moral relevance of ingroup advantage. Manuscript 1 examined whether a male or female confronter would be more accepted by men and whether one of them would elicit more positive privilege attitudes. We considered moral and image appraisal as potential mechanisms, explaining positive and negative reactions to the confrontation, respectively. Manuscript 2 investigated whether a contextualized style of confrontation that

[2] FINTA (female, intersexual, non-binary, transsexual, and agender) workers often provide unpaid care work in this scenario, which awards others (i.e., men) with the time and opportunity to do paid work, participate in politics or experience self-expression and leisure (Haug, 2008).

connects White privilege to Germany's involvement in colonialism can increase privilege acceptance and political solidarity by activating social justice morality.

Both manuscripts were inspired by the (shaky evidence of the) 3D model. As such they both tested parts of the model's hypotheses and illuminated the different mechanisms involved in responses to privilege confrontations. Specifically, Manuscript 1 examined the 3D model in the context of gender inequality. Accordingly, we tested whether a confrontation with male privilege would induce meritocratic threat and group-image threat, and whether these are respectively negative and positive mediators for privilege attitudes. We failed to find empirical support for these hypotheses and demonstrated that our alternative account rooted in the distinction between image and moral appraisal of male privilege is better equipped to predict privilege attitudes. Furthermore, we found that derogation toward the confronter is an underexplored response to privilege confrontations; a potential fourth "d". In Manuscript 2, we started by attempting to replicate the role of life hardships as a defensive response to White privilege confrontations. Adding to the somewhat thin evidence we did not find this effect across multiple studies and different sample demographics.

In the following sections, I communicate the research reported of the two manuscripts in more detail. A summary of the two manuscripts including publication status and authorship contribution can be found in Table 1.

Table 1. Manuscripts reported in this dissertation at a glance

	Manuscript 1: Appraisal of male privilege	**Manuscript 2: Addressing the "Why" in White privilege**
Research Questions	• Do men's responses to male privilege confrontations (especially backlash) depend on the confronter's gender? • In how far can the 3D model explain these responses, or is it more fruitful to distinguish between image and moral appraisal of privilege?	• Can we replicate the effect that confrontations with evidence of White privilege lead to claims of life hardships in German samples? • Can contextualizing White privilege in Germany's colonial history activate moral motives of social justice?
Total sample	*N* = 1463 male adults across three experiments (one student sample)	*N* = 646 White German adults across three experiments (one student sample)
Design	Mainly quantitative • 3 online experiments: manipulating gender of confronter. • Study 1 & 3 included a baseline control condition • Study 2 collected open answers to confrontation (thought protocol)	Quantitative • 3 online experiments: confrontation with evidence of White privilege versus baseline control • Study 2 & 3 added a contextualized confrontation condition
Main analyses	ANOVA, path analyses to probe for potential mediation, content coding of open answers	MANOVA, ANOVA, path analyses to probe for potential mediation
Main results	• Image appraisal had, contrary to the 3D model, a *negative* effect on privilege attitudes. • Moral appraisal explained positive reactions to the confrontation. • The female confronter was more threatening on both image and moral dimension than the male confronter, explaining lack of overall effects on privilege attitudes. • The female confronter was derogated more strongly than the male confronter even when evaluation was not prompted (i.e., open answers format).	• Effect of evidence confrontations on life hardships could not be replicated across three experiments. • Life hardships were only inconsistently related to privilege acceptance. • The contextualized confrontation increased moral motives of social justice, but only among White German students and White Germans with a migration background. • Social justice moral motives may mediate between confrontation and privilege acceptance as well as intentions to act in political solidarity with Black people.
Main implications	• 3D model may need substantial revision: group-image threat seems to be a negative influence, moral threat should be added as a positive influence, and derogation is an important additional outcome. • Privilege interventions are more effective and produce less backlash when they avoid triggering image concerns.	• Casts doubt about the alleged primary role of meritocratic threat in response to confrontations with White privilege. Effect might be culturally specific to the US (and UK). • Leveraging negative collective memory *can* increase the effectiveness of White privilege interventions among Germans who are more critically attached to their national identity. Induction of a moral redemption or moral growth mindset is a potential mechanisms.
Publication status	Published in *Journal of Experimental Social Psychology* doi: 10.1016/j.jesp.2023.104492	Submitted to *European Journal of Social Psychology* (May 28, 2023)
Authorship contribution and main role of co-authors	Eckerle, F.: 70% Rothers, A.: 15% (co-developing idea) Kutlaca, M.: 6% (close feedback to the manuscript and Study 3) Henss, L.: 3% (Study 1 was MA thesis) Agunyego, W.: 3% (Study 1 was BA thesis) Cohrs, J. C.: 3% (conceptual feedback)	Eckerle, F.: 80% Lienen, C. S.: 10% (co-developing idea, minor writing on collective memory) Kutlaca, M.: 4% (close feedback to the manuscript) Kotzur, P. F.: 3% (conceptual feedback) Cohrs, J. C.: 3% (conceptual feedback)

Manuscript 1: Appraisal of male privilege: On the dual role of identity threat and shame in response to confrontations with male privilege

The ingroup/outgroup positionality of confronters has been studied extensively in research on sexism confronters with equivocal results: While some research finds that male allies are more effective confronters of sexist behavior, other research finds that male allies trigger more moral concerns in fellow men, leading to stronger derogation – a black sheep effect (Kutlaca, Becker, & Radke, 2020; Marques & Paez, 1994). In Manuscript 1, my collaborators and I investigated how confronter gender affects responses to confrontations with male privilege. We started with an investigation whether threats to the ingroup's image would be associated with constructive attitudes (as the 3D model would suggest) or defensive attitudes (as research on confrontations with ingroup transgressions would suggest) toward ingroup privilege. Following this, we examined different responses to a male privilege confrontation. We then examined the role of image appraisal (image threat and image) and moral appraisal (moral threat and moral shame) to explain these outcomes.

We conducted three preregistered online experiments with adult male community samples (N = 1463). To confront with male privilege, we used an adapted article published by an influential German newspaper entitled "Ja, alle Männer" in all three studies (Dörr, 2019). The shortened article addressed male privilege in terms of gender pay gap, overrepresentation in leadership roles and in everyday situations such as walking home alone at night. To manipulate confronter positionality, we changed the prepositions and author name to indicate that the article was written by a male author or kept the article in its original format to indicate that the article was written by a female author. Our expectations regarding the effect of confronter gender were originally based on the black sheep effect, and thus we expected a stronger derogation of the male confronter. We also expected that the male confronter would elicit more moral threat and the female confronter would elicit more image threat, explaining more positive reactions to the male versus female confronter. In two of the three experiments, we also used a control group to

measure the baseline reaction to confrontations with gender inequality (without mentioning privilege) by summarizing results of a recent OECD report on gender inequality (OECD, 2017). Our hypotheses regarding the 3D model were tested by pooling both confrontations and comparing them against this control group. In accordance with the model, we expected a positive effect through group-image threat and a negative effect through meritocratic threat.

Over the course of the three studies, we found meager support for the 3D model: Meritocratic threat was not affected by the manipulation and group-image threat was *negatively* related to various measures of privilege acceptance as well as derogation of the confronter (Study 1). Instead, results supported our new appraisal-based model. Specifically, we found that image appraisal (image threat -> image shame) explained negative reactions toward the confrontation and that moral appraisal (moral threat -> moral shame) explained positive reactions (Study 3). However, contrary to our expectations, the female confronter did not specifically trigger image appraisal and the male confronter did not specifically trigger moral appraisal. In contrast, the female confronter was perceived to be more threatening on both the image and moral dimension, which led to an overall null effect between both confronters regarding privilege attitudes. Contrary to the black sheep effect, the female confronter was derogated more strongly in all three experiments, which was partly explained through higher levels of image appraisal in the female confronter condition. Additionally, the information on gender inequality (control condition) led to slightly more positive reactions than both confrontations.

We also coded texts written by participants as a thought protocol while reading the confrontation (Study 2). This was done to gauge which are actually relevant responses to the confrontation. The categories we identified were *accepting privilege and calling for things to change, justifying privilege and personal status, rejecting privilege and gender inequality,* and *derogating the confronter and feminism*. These categories could be well interpreted in relation to the three d's

proposed in the 3D model (with the addition of derogation) and were used to devise bipolar scales to measure privilege attitudes with less bias in the final study.

In summary, the results of Manuscript 1 challenge the validity of the 3D model and provide initial evidence based upon which we proposed the following adaptations: Group-image threat should be considered as a negative mediator. Moral threat should be included into the model as a positive mediator. Derogation or backlash should be added as an important fourth response to privilege confrontations.

Manuscript 2: Addressing the "Why" in White privilege: The limits of mere evidence confrontations and the potential of contextualizing White privilege in national colonial history

In Manuscript 2, we investigated the role of life hardships in response to a mere evidence confrontation, which has been used frequently in previous research. Additionally, we argued that a contextualized confrontation of White privilege might motivate reflection on the immorality of the status quo and would thus lead to privilege acceptance and political solidarity. Contextualization is understood as providing an account of privilege, which "outlines limits of individual responsibility" (Bergkamp, Martin, & Olson, 2022, p. 34). We expected that communicating White privilege in this way would reduce the tendency to individualize privilege and provide an opportunity for moral redemption of the ingroup through morally engaging in social justice. To test this, we examined how contextualizing White privilege in Germany's involvement in colonialism affects social justice morality and whether social justice morality explains privilege acceptance and solidarity-based collective action. We also considered boundary conditions to this effect, such as low preference for historical closure. This was chosen, because advocating historical closure (e.g., letting the past rest) indicates a defensive stance toward negative collective memory, which the contextualized confrontation taps into.

The manipulation material for the confrontation with mere evidence of White privilege as well as the main dependent variables to measure privilege acceptance were adapted from two previous

publications that found the effect on life hardships (Murdoch & McAloney-Kocaman, 2019; L. T. Phillips & Lowery, 2015). The contextualized confrontation extended this confrontation by prefacing a short text on Germany's involvement in colonialism, focusing on how the extraction of work force and resources has contributed to the emergence of White privilege. We also added a control condition in each experiment to avoid an overestimation of effects due to higher engagement in the experimental condition (Lane et al., 2021).

Across all three online experiments (N = 646), two of which were preregistered, we could not replicate the effect that confronting White people with evidence of White privilege would lead them to claim more life hardships, even though the statistical power was comparable or higher than in the original studies. Life hardships were also only inconsistently related to privilege acceptance and solidarity-based collective action intentions. In total, the mere evidence confrontation neither triggered accepting nor defensive responses.

The effectiveness of the contextualized confrontation was tested in parallel in the two latter experiments and results were promising but somewhat mixed: In a student sample (Experiment 2), we found the positive effect of the contextualized confrontation versus a control condition as hypothesized. However, there was no effect in a broader adult community sample (Experiment 3). One possible explanation was the strong representation of White participants with a migration background in that sample. Analyzing our model separately in a sample with and without migration background revealed that White Germans with a migration background exhibited a similar pattern as White students did in Study 2, while White Germans without a migration background seemed less affected by the contextualized confrontation. Finally, we also found some evidence that preference for historical closure moderates the effect of the contextualized confrontation: Only people who tended to reject historical closure responded with more solidarity-based collective action intentions after the contextualized confrontation (compared to the control condition). There was no such effect in the mere evidence confrontation.

In sum, the three experiments add further doubt regarding the validity and generalizability of the life hardship effect, and they indicate that contextualizing White privilege in collective memory of colonialism *can* produce positive effects by triggering social justice morality. These positive effects seem to be limited to demographics that are more critically attached to their national identity.

Manuscript 1

Appraisal of male privilege: On the dual role of identity threat and shame in response to confrontations with male privilege

Frank Eckerle, Adrian Rothers, Maja Kutlaca, Larissa Henss, Whitney Agunyego, & J. Christopher Cohrs

Interventions that confront men with male privilege can threaten their social identity. Past research on White privilege confrontations has suggested group-image threat is a positive predictor of positive privilege attitudes. However, research on emotional appraisals of ingroup transgressions demonstrates that feelings of shame elicited by the ingroup's negative image lead to less constructive responses while feelings of shame elicited by the ingroup's moral failure motivate more positive reactions. To reconcile these different predictions, we examined how relevant dimensions of threat (to the ingroup's image or to moral values of the group) and subsequent experiences of shame (image shame, moral shame) affect men's privilege attitudes in response to male privilege confrontations. Across three preregistered experiments (N = 1463 men), we found that male confronters were evaluated more positively, which was explained through lower levels of image concerns towards male confronters. We also found evidence that morality concerns are related to more positive responses. Although the total effects on privilege attitudes were small, our findings highlight the importance for future research and practice to distinguish between the positive effects of moral threat through moral shame and the negative effects of group-image threat through image shame in privilege interventions.

Keywords: Male Privilege, Identity Threat, Morality, Shame, Backlash

Journal of Experimental Social Psychology 108 (2023) 104492

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Appraisal of male privilege: On the dual role of identity threat and shame in response to confrontations with male privilege☆

Frank Eckerle [a,*], Adrian Rothers [a], Maja Kutlaca [b], Larissa Henss [a], Whitney Agunyego [a], J. Christopher Cohrs [a]

[a] Philipps-University Marburg, Department of Psychology, Gutenbergstr. 18, 35032 Marburg, Germany
[b] Durham University, Department of Psychology, Upper Mountjoy, South Rd, Durham DH1 3LE, United Kingdom

ARTICLE INFO

Keywords:
Male privilege
Identity threat
Morality
Shame
Backlash

ABSTRACT

Interventions that confront men with male privilege can threaten their social identity. Past research on White privilege confrontations has suggested group-image threat is a positive predictor of positive privilege attitudes. However, research on emotional appraisals of ingroup transgressions demonstrates that feelings of shame elicited by the ingroup's negative image lead to less constructive responses while feelings of shame elicited by the ingroup's moral failure motivate more positive reactions. To reconcile these different predictions, we examined how relevant dimensions of threat (to the ingroup's image or to moral values of the group) and subsequent experiences of shame (image shame, moral shame) affect men's privilege attitudes in response to male privilege confrontations. Across three preregistered experiments (N = 1463 men), we found that male confronters were evaluated more positively, which was explained through lower levels of image concerns toward male confronters. We also found evidence that morality concerns are related to more positive responses. Although the total effects on privilege attitudes were small, our findings highlight the importance for future research and practice to distinguish between the positive effects of moral threat through moral shame and the negative effects of group-image threat through image shame in privilege interventions.

Gender inequality and sexism are often framed in terms of disadvantages for women. However, feminist scholars point out the importance of also acknowledging and dismantling male privileges to overcome sexism, patriarchal arrangements, and masculinist standards (Galea & Gaweda, 2018; Pease, 2010). Past research has shown, for example, that becoming aware of one's group-based privileges and perceiving them to be unjust is a central component in motivating political allyship against systems of inequality (Droogendyk, Wright, Lubensky, & Louis, 2016; Kutlaca, Radke, Iyer, & Becker, 2020). These group-based privileges are unearned (McIntosh, 1988), automatically awarded (Barajas, 2022), often normalized and, therefore, invisible advantages based on group-membership (Knowles, Lowery, Chow, & Unzueta, 2014).

Although events that raise men's awareness of male privilege can increase their willingness to reflect on their sense of entitlement, this can also backfire, leading to reactionary commentaries (e.g., Fortgang, 2014), counter-movements (Coston & Kimmel, 2013), and, more generally, backlash (Flood, Dragiewicz, & Pease, 2021; Pease, 2021). As one example, the #NotAllMen hashtag frequently appeared in social media after public discussions of incidents of gender-based violence, rape, and murder, reframing discussions about the cultural dominance of misogyny (Jones, Chałupnik, Mackenzie, & Mullany, 2022; Maricourt & Burrell, 2022). As Nicholas and Agius (2018) conclude, "[t]hese contemporary backlashes naturalize essentialized gender hierarchies and their intersection with other naturalized hierarchies to claim that feminism has gone too far" (p. 31). Outcomes of privilege interventions also span a broad range from defensive claims of "reverse discrimination" (Coston & Kimmel, 2013; Norton & Sommers, 2011) over null-effects (Puryear, Steele, Lawler, & Vandello, 2019) to increased awareness about privilege and the underlying system of inequality (Ancis & Szymanski, 2001; Case, Hensley, & Anderson, 2014). This highlights the need to examine potential mechanisms behind positive

☆ This paper has been recommended for acceptance by Evava Pietri.

* Corresponding author at: Gutenbergstr. 18, FB04 Psychologie, 35032 Marburg, Germany.
E-mail addresses: frank.eckerle@uni-marburg.de (F. Eckerle), rothersa@uni-marburg.de (A. Rothers), maja.kutlaca@durham.ac.uk (M. Kutlaca), henssla@uni-marburg.de (L. Henss), agunyegw@uni-marburg.de (W. Agunyego), christopher.cohrs@uni-marburg.de (J.C. Cohrs).

https://doi.org/10.1016/j.jesp.2023.104492
Received 21 October 2022; Received in revised form 17 May 2023; Accepted 19 May 2023
Available online 27 May 2023

and negative responses to privilege interventions. In the present paper, we investigate the role of different dimensions of identity threat and shame, focusing on the question of whether reactions to privilege confrontations can be explained by appraisals of the confrontation that are rooted in either concern for the ingroup's image or morality. We draw from two separate strands of research to predict reactions toward confrontations with ingroup privilege: Research on privilege confrontations and research on confrontations with ingroup transgressions and moral failure. We summarize the role these research strands assign to identity threat, highlight contradictions, and then propose an integrated appraisal-based model.

1. Identity threat in privilege confrontations: the 3D model

The 3D model (Knowles et al., 2014) provides a theoretical framework to explain why White people, on the one hand, *deny*, *distance* or, on the other hand, *dismantle* their White privilege. According to this model, reactions toward privilege confrontations depend on the extent to which the confrontation elicits meritocratic threat ("a threat to the personal self and its competence"; Knowles et al., 2014, p. 598) and/or group-image threat ("a threat to the collective self and its moral connotations"; Knowles et al., 2014, p. 598). While meritocratic threat is expected to motivate denying or distancing from privilege in order to protect the agentic integrity of the self (Steele, 1988), a threat to the group's image should have more positive effects. Hence, according to the 3D model, to elicit constructive responses, confrontations should ideally threaten the ingroup's image without threatening the meritocratic belief system.

The current research tests both threats suggested by the 3D model in a single study. However, we don't claim to test the model's overall validity, because the 3D model was conceived to explain reactions to White privilege and we are focusing on reactions to confrontations with male privilege; even though the authors also cite papers on male privilege (e.g., Branscombe, 1998) and apply them to White privilege; Knowles et al., 2014, p. 599). Although there are good arguments to commensurate research on the psychology of privilege (Nixon, 2019; Schmitt, Behner, Montada, Müller, & Müller-Fohrbrodt, 2000), there are also important domain differences: For example, confrontations with gender bias are perceived as less threatening than confrontations with racial bias (Czopp & Monteith, 2003). However, a comparison between reactions to different domains of privilege is beyond the scope of this paper. Our focus lies instead on disentangling the role of identity threat more clearly.

2. The role of morality and shame in confrontations with ingroup transgressions

Research on the effects of confrontations with moral failure and ingroup transgressions has investigated the role of identity threat based on the emotional appraisal of the confrontation, finding that the emotional appraisal of such a confrontation can predict subsequent reactions (Gausel, Leach, Vignoles, & Brown, 2012). When moral failure is perceived to be a specific self-defect, this can lead to feelings of shame, and subsequently to self- and social improvement. However, if moral failure is perceived as a condemnation by others, this more likely leads to feelings of rejection and self-defense (Cibich, Woodyatt, & Wenzel, 2016; Gausel et al., 2012). Relatedly, Allpress, Brown, Giner-Sorolla, Deonna, and Teroni (2014) revealed that highlighting transgressions of an advantaged ingroup can either trigger image shame (based on a concern for the image of the group) or moral shame (based on undermined moral essence of the group). Image shame motivated participants to maintain the ingroup's positive image in the eyes of others and led to rejection and negative attitudes toward the outgroup. Moral shame motivated restoration of damaged ingroup morality. Thus, image shame explained negative and moral shame positive outgroup orientations.

On the one hand, the 3D model expects image concerns to be a predictor of positive reactions to privilege confrontations. On the other hand, research has found that image shame predicts more negative intergroup attitudes. The key to resolve this contradiction might lie in the 3D model's definition of group-image threat, which does not differentiate between threats to the ingroup's image and threats to the "moral standing" (Knowles et al., 2014, p. 602) of the group. To clarify this, we argue that group-image threat and moral threat are separate constructs. Group-image threat is rooted in a concern that others might negatively evaluate the ingroup (see also Shuman, Johnson, Saguy, & Halperin, 2018). Moral threat, instead, is rooted in a threat to the moral essence of a group (see also Allpress et al., 2014). We argue that examining moral threat independent from image threat is important, because regulating morality seems to be a core function of social identities (Ellemers, Spears, & Doosje, 2002; Leach, Bilali, & Pagliaro, 2015). Past research found, for example, that concerns about negative ingroup evaluation led to *negative* intergroup outcomes, while concerns about failing to meet the moral standards of the ingroup have been linked to *positive* intergroup outcomes (Ellemers, Pagliaro, & Barreto, 2013; Ellemers & van den Bos, 2012). As such, threats to the social image likely result in the condemnation of others, whereas threats to the moral self can motivate self-improvement (Gausel & Leach, 2011) and pro-social behavior (Gausel, Vignoles, & Leach, 2016). Whether a confrontation is perceived to be a threat to the ingroup's image or a violation of moral values of the ingroup might therefore explain subsequent responses to the confrontation. Specifically, we expect that an image appraisal path (*group-image threat* –> *image shame*) explains negative and a moral appraisal path (*moral threat* –> *moral shame*) explains positive response to privilege confrontation (see Fig. 1 for an overview of the theoretical model).

3. The role of ingroup/outgroup confronter status on privilege appraisal and responses to their confrontation

The potential existence of two distinct paths or modes of appraising ingroup threat raises another important question: *When* do men perceive a male privilege confrontation as a threat to their group's image versus a threat to their ingroup's moral integrity? Previous research found, for example, that when ingroup members (compared to outgroup members) question ingroup morality, they trigger more emotional and physiological reactions (Ellemers et al., 2013). We suggest that our appraisal-based approach is able to explain potential differences in men's reactions to a male privilege confrontation depending on the gender identity of the confronter.

Research on sexism confronters found that devaluation of the confronter can serve as a defense strategy against the accusation of sexism. Men perceive other men who confront sexist behavior to be more legitimate than women who confront the same behavior (Drury & Kaiser, 2014). Men are also more likely to expect women who confront others with sexist behavior to act out of self-interest, which might explain their reduced legitimacy (Czopp & Monteith, 2003; Glick, 2014), and why confronters from the advantaged ingroup are perceived to be more persuasive (Rasinski & Czopp, 2010). While the role of backlash has not been examined in the context of privilege confrontations, this literature would suggest less backlash against male (compared to female) confronters.

More recent research revealed that men are more likely to lash out against fellow men (compared to women) who confront sexist behavior (Kutlaca, Becker, & Radke, 2020). The negative evaluation of male (ingroup) confronters can be interpreted as a strategy to maintain positive group esteem by derogating deviating group members – the so-called Black Sheep Effect (Marques, Abrams, Paez, & Martinez-Taboada, 1998; Marques & Paez, 1994; Pinto, Marques, Levine, & Abrams, 2010). This is in line with findings that ingroup members are more likely to threaten the ingroup's image (Brambilla, Sacchi, Pagliaro, & Ellemers, 2013). As the logic of the Black Sheep Effect seemed closest to the 3D model's postulations, we originally expected that male

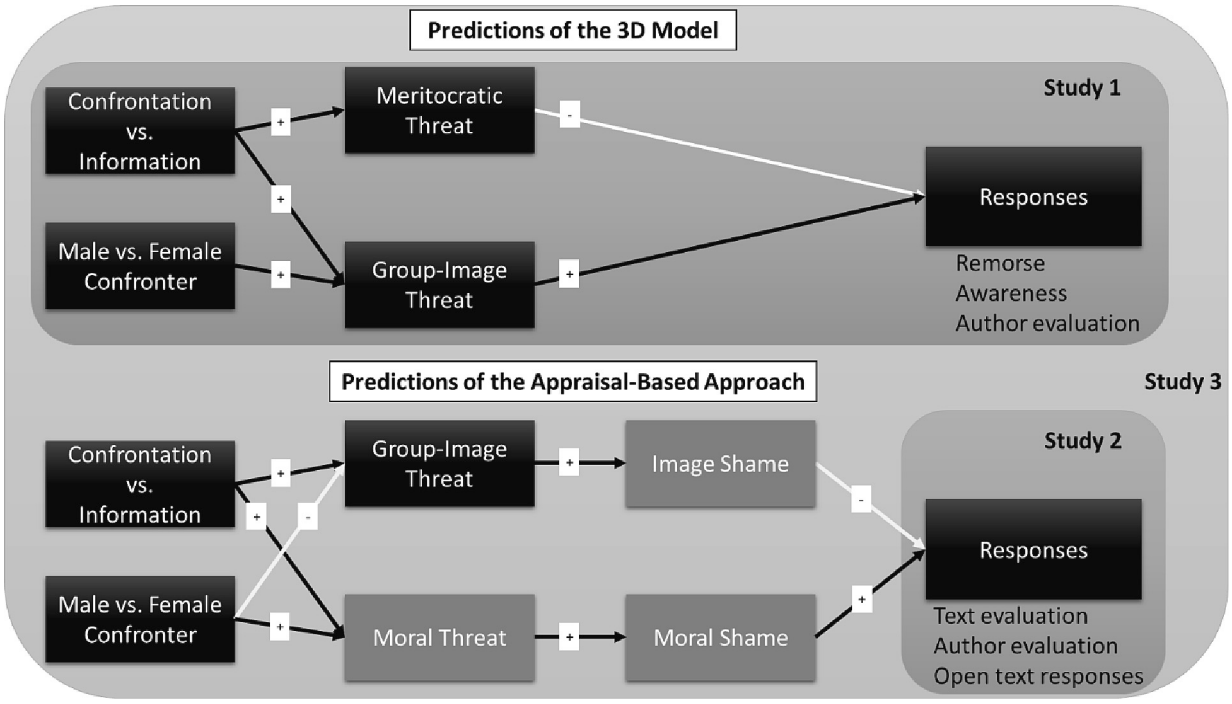


Fig. 1. Overview of the theoretical model and the three studies.
Note. White arrows represent negative expected effects of each model. Grey boxes represent concepts borrowed from literature on moral failure and ingroup transgressions. Moral threat was implied to be a part of group-image threat in the 3D model, but past research only measured the aspect of image concern. Study 3 used the same outcomes as Study 1 and 2 with the exception of privilege remorse. Study 3 also provided three new outcomes based on open responses in Study 2.

confronters receive more backlash.

Confronter evaluations were not considered in the 3D model. We believe that the relation between derogation and the three key outcomes of 3D model depends on the individual's response to the confrontation. For example, derogation may be used to discredit a feminist discourse, which brings it closer to denial, or it may be used to defend one's self-esteem and thus serve as a distancing strategy. Clarifying the exact role of confronter evaluations in relation the 3D model is beyond the scope of the current paper. However, correlations between these outcomes might provide some initial directions.

4. Overview of the model and studies of this paper

The current paper develops an appraisal-based model to (male) privilege confrontations across three studies: In Study 1, we investigated the explanatory value of group-image and meritocratic threat as proposed in the 3D model. In Study 2, we examined the role of shame as a mechanism behind the reactions to privilege confrontations and compared text-based responses toward a male and female confronter. In Study 3, we tested the full appraisal-based model, examining the role of the moral appraisal and image appraisal paths in explaining reactions to privilege confrontations and confronters. Study 3 also provides a conceptual replication of Study 1. To evaluate the specific effect of a *confrontation*, we used material that directly addresses the reader ("you men" / "we men") and compared it to a non-confrontative *information* on the current state of gender inequality. Hence, this control condition provides a baseline with a similar context but without attributions of responsibility or calls for action and should therefore be perceived as less threatening. Furthermore, we compared the effects of confronter gender in all three studies. An overview over the two theoretical models and which study tests which part of these models is presented in Fig. 1.

Given the at times contradictory state of the relevant bodies of literature, conceptual replications across multiple studies are needed to confirm results and to gradually reveal underlying mechanisms. Therefore, the three studies were designed incrementally. To provide transparency about changes of our hypotheses between studies, we preregistered each study based on the results of the previous one. We link to the pre-registrations in the Open Practices section below. All measures, manipulations, and exclusions are reported in the method section of each study and the supplementary material is referenced for additional details. All *t*-tests are based on Welch's statistic. In all studies, sample size was determined before any data analysis.

5. Study 1

In this study, we aimed to test the 3D model's suggestion that meritocratic and group-image threat are oppositional mediators explaining less and more positive responses to the privilege confrontation respectively. We also explored whether threat is affected by the confronter's gender. The hypotheses of Study 1 were preregistered based on the 3D model's propositions and are presented here in a summarized form:

Hypothesis 1. A confrontation with male privilege (compared to a low-threatening information) will affect attitudes toward male privilege (privilege awareness and remorse) to the degree that it triggers group-image threat (a positive mediator) and meritocratic threat (a negative mediator).

Hypothesis 2. The level of group-image threat and meritocratic threat will be different when the confronter is male versus female and both threats will explain subsequent positive or negative effects on attitudes toward male privilege as in H1.

Hypothesis 3. The male confronter will be evaluated more negatively than the female confronter.

Exploratorily, we also considered an alternative setup based on our appraisal-based approach. For this, privilege remorse was used as a proxy for the moral appraisal path parallel to group-image threat, because the privilege remorse subscale has some overlap with moral shame items (see Allpress et al., 2014; Pinterits, Poteat, & Spanierman,

2009). A positive effect of privilege remorse on privilege awareness and a negative effect of group-image threat on privilege awareness would be in line with previous research on the appraisal of moral failure (Gausel et al., 2016; e.g., Gausel & Leach, 2011). The effect should also be stable when privilege awareness is regressed on both variables simultaneously. This exploratory approach was not preregistered and appears in the additional analyses section below.

5.1. Method

5.1.1. Participants

We collected data through an online questionnaire hosted on SosciSurvey (Leiner, 2019a), which we distributed via the mailing lists of student body faculties of different German universities and snowballing on social networks. $N = 1121$ participated in the survey of which $n = 775$ completed the survey and identified as male. We excluded participants based on the following preregistered criteria: $n = 8$ were excluded because they had >10% missing data and $n = 118$ were excluded because they incorrectly answered an attention check, asking about the author of the text (male or female confronter or OECD information). The final sample consisted of $N = 654$ male participants with a mean age of 25.90 ($SD = 6.91$) years. Psychology students were offered course credits for their participation.

5.1.2. Materials & procedure

Unless noted otherwise, all dependent variables were measured on a 7-point scale from 1 (strongly disagree) to 7 (strongly agree). Correlations, means and standard deviations of the main dependent variables appear in Table 1.

Demographics. Participants provided demographic data on their age, gender, and education at the beginning of the survey.

Manipulation of confrontation and gender. Next, we randomly assigned participants to one of three conditions: female confronter, male confronter, and information. In both confronter conditions, participants read an adapted version of an article about male privileges that was published in a leading German newspaper (Dörr, 2019). In the article, the author directly addressed the reader, either as "you" or "we", implying their own gender. Additionally, the name of the alleged author was given a female name (Johanna) in the female confronter condition or a male name (Johannes) in the male confronter condition. In the information condition, participants instead read a text that summarized statistical data from the OECD regarding gender inequality around the same issues that the newspaper article focused on, such as gender pay gap and domestic violence (OECD, 2017). The article mentioned an OECD work group as the source of this information and there was no gender attached to the authorship in the information condition.

Author evaluation. We asked participants how positively they perceived the author of the text they had just read on a feeling thermometer ranging from 0 (very negative) to 100 (very positive). We also measured ratings on social judgement dimensions, which are reported in the supplementary materials (section A).

Group-image threat. Group-image threat (Shuman et al., 2018) was measured with two items: "While I was reading the text, I had the impression that men look bad because of their privilege" and "While I was reading the text, I had the impression that advantaging men makes other people have a negative image of my gender" ($\rho = 0.79$).

Meritocratic threat. We created two items to tap into meritocratic threat: "It makes me angry when feminists think that my successes are only due to my gender" and "It is not okay when feminists think that my good social status is only due to my gender" ($\rho = 0.84$).

Attitudes toward male privilege. Male privilege awareness ($\alpha = 0.81$) and male privilege remorse ($\alpha = 0.87$) were each measured with three items using the respective subscales of the White Privilege Attitudes Scale (Pinterits et al., 2009) adapted to male privilege. Example items are: "I am ashamed that the system is stacked in my favor because I am male" (remorse) and "Men have it easier than women" (awareness).

Attention check and additional variables. As an attention check, we asked participants to indicate the gender of the author (male, female, work group) at the end of the survey. Some additional variables for research questions beyond the scope of the current paper have also been collected. See supplementary materials section A for more details.

5.2. Results

5.2.1. Test of the 3D model and of the effect of confronter gender

A path model was computed using the pathj module (Gallucci, 2021) for jamovi (The jamovi project, 2021) with maximum likelihood estimation and robust standard errors. Group-image threat and meritocratic threat were regressed on the exogenous factor condition and privilege awareness and privilege remorse were regressed on both threats. All continuous variables were mean-centered. The condition variable was difference-coded to represent the two relevant contrasts (information-confrontations and male-female confronter). We also estimated the covariance between group-image and meritocratic threat. The path model had excellent fit, $\chi^2(4) = 3.18$, $p = .57$, CFI = 1.00, RMSEA = 0.00 ($p = .93$, 95% CI [0.00 0.05]), SRMR = 0.01.[1] Table 2 provides an overview of all indirect effects (IE) tested and Fig. 2 displays the results graphically.

Group-image threat did neither affect privilege awareness, $b = 0.01$, 95% CI [−0.06 0.08], $\beta = 0.01$, $p = .78$, nor privilege remorse, $b = 0.07$, 95% CI [−0.004 0.13], $\beta = 0.07$, $p = .07$, but was related to more negative author evaluations, $b = -1.87$, 95% CI [−3.16–0.58], $\beta = -0.12$, $p = .005$. Meritocratic threat was negatively related to all three outcomes: privilege awareness, $b = -0.42$, 95% CI [−0.48–0.36], $\beta = -0.55$, $p < .001$, privilege remorse, $b = -0.27$, 95% CI [−0.33–0.21], $\beta = -0.35$, $p < .001$, and author evaluation, $b = -4.82$, 95% CI [−5.95–3.70], $\beta = -0.35$, $p < .001$. Thus, even though zero-order correlations between group-image threat and privilege attitudes were significant, this effect disappeared when including meritocratic threat in the model. Therefore, Hypothesis 1 was only confirmed regarding the

Table 1
Correlation matrix with relevant dependent variables.

	1	2	3	4	5
Group-Image Threat	4.07 (1.62)				
Meritocratic Threat	0.38***	4.40 (1.89)			
Privilege Awareness	−0.20***	−0.54***	4.83 (1.46)		
Privilege Remorse	−0.05	−0.33***	0.46***	2.46 (1.46)	
Author Evaluation	−0.30***	−0.39***	0.47***	0.37***	49.34 (26.40)

Note. *** p < .001. An extended table with additional variables appears in the supplementary materials (section A). Means (and standard deviation) appear on the diagonal.

[1] Note that the order of measured variables in mediation analyses does not imply a causal order (Spencer, Zanna, & Fong, 2005). In our case, past experimental research provides some confidence that identity threat in fact *causes* attitude change (Branscombe, Ellemers, Spears, & Doosje, 1999; Steele, Spencer, & Aronson, 2002). Nevertheless, we want to emphasize the correlational nature of the associations between the variables image threat / moral threat, response to privilege confrontation and author evaluation in our studies. Another concern about mediation analyses is that mediator variables often have large overlap with the predictor or outcome variable. Conceptually, we see little overlap both between our mediator (identity threat) and the independent variable (confronter gender) and between the mediator and the dependent variables (author/confronter evaluation and response to privilege confrontation). Correspondingly, correlations between the mediator and dependent variables were moderate or small, $r < |0.30|$ (see supplementary materials).

Table 2
Overview of indirect effects.

	Mediator	Outcome	*b*	LCI	HCI	β	*p*
Indirect effects for the contrast Male-Female							
IE1	Image	Awareness	0.00	−0.03	0.03	−0.001	0.781
IE2	Image	Remorse	−0.03	−0.06	0.01	−0.008	0.112
IE3	Image	Evaluation	0.79	−0.01	1.58	0.012	0.052
IE4	Merit	Awareness	−0.05	−0.20	0.10	−0.014	0.507
IE5	Merit	Remorse	−0.03	−0.13	0.07	−0.009	0.508
IE6	Merit	Evaluation	−0.59	−2.33	1.16	−0.009	0.508
Indirect effects for the contrast Information-Confrontations							
IE7	Image	Awareness	−0.01	−0.06	0.05	−0.003	0.780
IE8	Image	Remorse	−0.05	−0.11	0.00	−0.016	0.072
IE9	Image	Evaluation	1.46	0.31	2.60	0.026	0.013
IE10	Merit	Awareness	0.12	−0.01	0.25	0.039	0.074
IE11	Merit	Remorse	0.08	−0.01	0.16	0.025	0.079
IE12	Merit	Evaluation	1.37	−0.15	2.88	0.024	0.077

Note. IE1 to IE12 = labels of indirect effects.

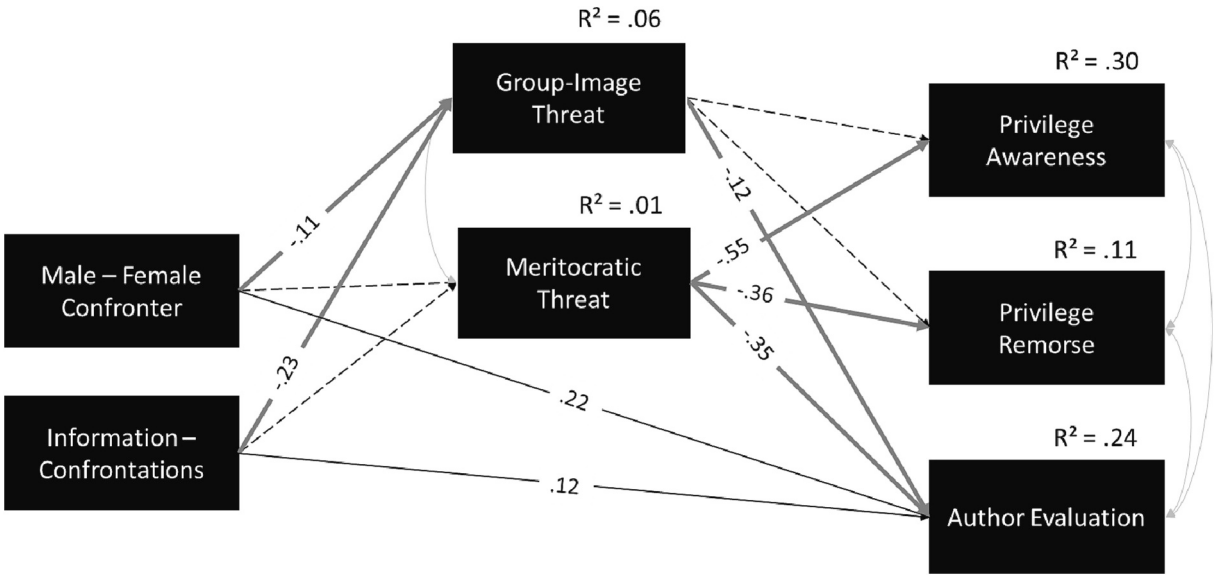


Fig. 2. Path model with group-image threat and meritocratic threat as mediators.
Note. Estimates are standardized. Direct effects appear as thin arrows. Non-significant paths appear dashed ($p > .05$).

negative effect of meritocratic threat on male privilege attitudes.

The information-confrontations contrast indicated that participants in the confrontation conditions perceived higher levels of group-image threat, $b = -0.78$, 95% CI [−1.04–0.51], $\beta = -0.23$, $p < .001$, but meritocratic threat was not significantly affected, $b = -0.28$, 95% CI [−0.59 0.03], $\beta = -0.07$, $p = .07$. Author evaluation was more positive in the information condition, $b = 6.75$, 95% CI [3.16 10.33], $\beta = 0.12$, $p < .001$. In this contrast, only the indirect effect through group-image threat on author evaluation was significant (IE9). This pattern speaks against Hypothesis 1, as group-image threat seems more likely to be a predictor of backlash.

Regarding the male-female confronter contrast, the male confronter triggered less group-image threat, $b = -0.40$, 95% CI [−0.711–0.13], $\beta = -0.11$, $p = .005$, but there was no effect on meritocratic threat, $b = 0.12$, 95% CI [−0.24 0.48], $\beta = 0.03$, $p = .51$. The indirect effects of confronter gender on privilege awareness (IE1, IE4) and remorse (IE2, IE5) were all non-significant. Therefore, Hypothesis 2 was also only partly supported. Contrary to Hypothesis 3, the male confronter was evaluated more positively, $b = 14.05$, 95% CI [10.02 18.08], $\beta = 0.22$, $p < .001$.

We conducted a sensitivity power analysis to estimate the minimum indirect effect obtainable given the standard criteria (1-β = 0.80, α = 0.05) and sample size (N = 654). For this, we used Monte Carlo simulations with 1000 replications, 20,000 draws per replication, and a confidence level of 95%, requesting statistical power given standardized model coefficients (Schoemann, Boulton, & Short, 2017). In the first step, we calculated power based on the standardized coefficients of a medium sized indirect effect (IE8). Statistical power to find this indirect effect was 0.52. Next, we increased the coefficients of both paths by the same value (using increments of 0.001) and ran the analysis again until the target power of 1-β = 0.80 was reached. According to this procedure, the minimum indirect effect (β_a*β_b) obtainable under standard criteria and $N = 654$ was $|\beta_{ab}| = 0.029$.

5.2.2. Additional analyses

To test our appraisal-based approach, linear regressions with privilege remorse and group-image threat as predictors of privilege awareness and confronter evaluation were conducted. The predictors explained $R^2_{adj} = 0.24$ of the variance of privilege awareness and $R^2_{adj} = 0.21$ of variance of author evaluation. Results supported the notion that privilege remorse (as a proxy of moral shame) would be positively related to privilege awareness, $b = 0.46$, 95% CI [0.38 0.51], $\beta = 0.45$, $p < .001$, while group-image threat would be negatively related to privilege awareness, $b = -0.16$, 95% CI [−0.24–0.10], $\beta = -0.17$, $p < .001$. We found the same pattern regarding author evaluation for privilege remorse, $b = 6.37$, 95% CI [0.28 0.42], $\beta = 0.35$, $p < .001$, and group-

image threat, $b = -4.49$, 95% CI [−0.35–0.21], $\beta = -0.28$, $p < .001$.

5.3. Discussion

Study 1 provides mixed support and some counterevidence for the hypotheses derived from the 3D model. Regarding Hypothesis 1, we found that group-image threat was *negatively* related to privilege attitudes, but this effect disappeared when investigated in the context of meritocratic threat. Meritocratic threat was, as expected, negatively related to both privilege awareness and remorse. Although, as predicted, the confrontation (compared to the information) increased levels of group-image threat, we found no effect of the confrontation on meritocratic threat. Furthermore, the indirect effects through both threats on privilege attitudes were non-significant. Regarding Hypothesis 2, we found that the male confronter triggered less group-image threat while there was no difference in meritocratic threat. Opposite to Hypothesis 3, the male confronter was evaluated *more positively* than the female confronter. Lower levels of group-image threat in the male confronter condition seemed to be a relevant mechanism. This finding contradicts the assumptions based on the Black Sheep Effect and 3D model. It is more in line with previous research on sexism confrontations which found that female confronters receive more negative judgments than male confronters (Czopp & Monteith, 2003; Eliezer & Major, 2012) and that ingroup confronters are in a better position to question the morality of an ingroup (Ellemers et al., 2013). The following studies therefore derive their predictions based on this literature.

Supporting our appraisal-based approach, additional analyses revealed that group-image threat was *negatively* and privilege remorse (as proxy of moral shame) *positively* related to privilege awareness. We found the same pattern for author evaluation.

6. Study 2

We conducted Study 2 to investigate whether men's responses to privilege confrontations differ as a function of the confronter's gender. We thus collected and content-coded text-based reactions. Based on the results of Study 1 and the previous finding that ingroup confronters are better positioned than outgroup confronters to question the morality of the ingroup (Ellemers et al., 2013), we expected that men have more positive privilege attitudes after reading a confronting text written by another man (vs. a woman; Hypothesis 1), that men have more positive attitudes toward a male (vs. female) confronter (Hypothesis 2), and that men react with less image concern toward a male (vs. female) confronter (Hypothesis 3). These hypotheses were preregistered.[2] We additionally preregistered an exploration of the effect on guilt and shame, expecting them to be indicators of threat. However, as feelings of guilt and shame cannot clearly distinguish between moral and image appraisal, which is the main interest of this manuscript, we report results on these measures in the supplementary materials (section B). Study 3 then focused more closely on separating moral and image appraisal.

6.1. Method

6.1.1. Participants

We circulated an online questionnaire via snowballing and within a course credit system (SONA) of a German university. We preregistered to collect $N = 250$ participants or stop data collection after four weeks. This was based on an a priori power analysis in G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) for a one-sided independent *t*-test with a target power of 0.80 and the targeted effect size $d = 0.30$, which was similar to the effect size of confronter gender on group-image threat in Study 1 ($d = 0.27$). The questionnaire was programmed using the same provider as in Study 1. Due to slow data collection, we extended the data collection period to six weeks, after which $N = 209$ participants completed the survey. In accordance with preregistered procedures, we excluded data from $n = 24$ participants because they indicated to already know the material used for manipulation, did not provide text in the thought protocol, did not identify as male, or because they had a time penalty above 75[3] (Leiner, 2019b). The final sample consisted of $N = 186$ male participants with a mean age of 25.52 ($SD = 7.49$) years and an above average subjective social status on a scale from 1 to 10 ($M = 6.16$, $SD = 1.61$). Psychology students were offered course credits for their participation.

6.1.2. Materials & procedure

Demographics. We collected data on participants' age, gender, and subjective social status on a 10-point scale with 1 = close to people at the top of society and 10 = close to people at the bottom of society (Hoebel, Müters, Kuntz, Lange, & Lampert, 2015) at the beginning of the survey. Subjective social status was reversed to simplify interpretation.

Manipulation. Participants were randomly allocated to either a male or female confronter condition. In both conditions, participants read the same text as in Study 1. The alleged first name of the author was "Daniel" in the male confronter condition and "Katharina" in the female confronter condition. We picked these names as they are strongly associated with the male and female gender, respectively, and are associated with comparable social characteristics in the German context (Nett, Dorrough, Jekel, & Glöckner, 2020). We placed an open text field next to each paragraph of text and provided the following instruction: "Below you will find a text excerpt from a popular German daily newspaper. To the right of the article are empty text fields that belong to the adjacent paragraphs. Please enter your thoughts, opinions, and associations to each paragraph. In research, this method is called 'thinking aloud'; we are therefore interested in everything that goes through your mind while reading. It is irrelevant, for example, whether a thought seems important or unimportant to you or whether it is pleasant or unpleasant. There is no right or wrong!".

Coding procedure and codebook. In our coding scheme we started with selecting potentially relevant strategies of neutralizing arguments from an existing classification (Kaptein & van Helvoort, 2019). We then read the texts and coded for the different strategies that appeared in the data. Afterwards, we sorted the codes into the different strategies of neutralization or different forms of acceptance. Finally, we summarized each code into one of four main dimensions of responses to privilege: accepting privilege and calling for social change, justifying privilege without directly denying gender-based inequality (e.g., legitimizing gender differences by naturalization), rejecting privilege (e.g., listing counterexamples of male disadvantage), and backlash against the confronter (e.g., victim blaming or devaluation). The codebook is summarized in Table 3.

The first and second author created the codebook after becoming familiar with recurring concepts in the data. Afterwards, the first author coded the data completely. To test inter-coder reliability, the second author coded 20% of answers blindly to the first coder's decisions. Inter-coder reliability was moderate on the code-level, Cohen's $\kappa = 0.57$, and substantial on the more inclusive categories (acceptance, justification,

[2] We also preregistered that men would trigger less image threat and planned to code for image threat in the open text responses (i.e., "The author makes men look bad"). However, participants' responses only seldomly allowed for a clear coding of image threat. Therefore, we did not test this hypothesis in this experiment and relied more strongly on the also pre-registered close coding of the text.

[3] We preregistered exclusion of participants at a relative speed index above 2 (see Leiner, 2019b). However, the relative speed index penalized participants who wrote short answers in the text fields compared to participants who were willing to elaborate their responses. As not providing data in the open text fields was also an exclusion criterion, we used the less restrictive penalty point variable for exclusion, which was also provided by SosciSurvey (DEG_TIME).

Table 3
Codebook overview.

Category	Code	Description
Acceptance	General acceptance	Acceptance of male privilege and self-reflection
	Change talk	Acceptance with reflections about changing attitudes or behavior or calling for social change
	Limited acceptance	Acceptance with some reservations or emotional ambivalence
	Supplemented acceptance	Acceptance and further reflections about other effects of male privilege, intersectionality, and complexity
Justification	Rejection of responsibility	Claiming personal hardship, attributing responsibility to others or other men, defending against accusation
	Powerless individual	Lack of personal influence or pointlessness of attempting change
	Legitimization	Legitimizing gender differences by naturalization, normativity, or other strategies
	Overgeneralization	Arguments beyond gender differences, pointers toward a universal compass of egalitarian values that diffuse the argument
Rejection	Single cases	Opposition to the systemic nature of inequality and privileging
	Mixed evidence	Allegation of "cherry picking" data, lack of evidence and counterexamples
	Reverse discrimination	Examples of alleged discrimination against men, allegation of female victim culture
	Whataboutism	Examples of other systems of inequality that are allegedly more important
	Unspecific rejection	Unspecified rejection (e.g., "bullshit")
Backlash	Victim blaming	Responsibility for inequality is attributed to women and their (bad) decisions
	Devaluation	Devaluation of the author and others with similar views, accusation of political division, being annoyed/angry at the discourse itself

rejection, backlash), Cohen's $\kappa = 0.63$ (O'Connor & Joffe, 2020).

Privilege attitude. We measured privilege attitude via one item that inquired to what degree participants strongly disagreed (0) to strongly agreed (100) with the opinion that was presented in the text. We chose this indirect measure to avoid effects of social desirability and response bias due to a transparent research interest.

Author evaluation. We measured author evaluation as in Study 1. We also measured difference in social judgement dimensions (competence, warmth, morality), which are reported in the supplementary materials (section B).

Additional measures. We provided a list of nine items measured on a scale from 1 (not at all) to 5 (very strongly) to measure possible emotional responses to the texts: shame, guilt, annoyance, anger, affliction, sadness, pride, gratitude, and confusion. Emotional reactions were explored in preparation for Study 3 and results are reported in the supplementary materials section B.

6.2. Results

Sensitivity power analysis via G*Power (Faul et al., 2007) returned $d = 0.37$ as the minimum effect size obtainable for one-sided independent t-tests based on 1-$\beta = 0.80$, $\alpha = 0.05$, and n of the female confronter ($n = 95$) and male confronter ($n = 91$) condition.

Each response category (acceptance, justification, rejection) was named by roughly two thirds of participants with rejection being the most frequent category (73.7% of participants). Based on the premise of our hypotheses (that a male confronter is more effective and more positively evaluated), we tested acceptance one-sided for the direction female confronter < male confronter and justification, rejection, and backlash one-sided for the direction female confronter > male confronter. Results indicate that men justified their male privilege less often toward a male confronter, $t(178.55) = -2.03$, $p = .02$, $d = -0.30$, and that a male confronter triggered less backlash responses, $t(177.53) = -1.87$, $p = .03$, $d = -0.27$. There was no difference in acceptance, $t(181.89) = 0.16$, $p = .56$, $d = 0.02$, and rejection, $t(183.75) = 0.51$, $p = .69$, $d = 0.07$. The relative frequency of each code is reported in the supplementary materials (section B).

Text evaluation were more positive in the male confronter ($M = 51.62$, $SD = 36.11$) than in the female confronter condition ($M = 46.42$, $SD = 34.08$) but contrary to Hypothesis 1 the effect was not significant, $t(180.73) = -1.01$ $p = .16$, $d = -0.15$. In line with Hypothesis 2 and results of Study 1 we found that the male confronter ($M = 46.52$, $SD = 26.96$) was more positively evaluated than the female confronter ($M = 39.92$, $SD = 24.54$), $t(180.13) = -1.74$, $p = .04$, $d = -0.26$. Text and author evaluation correlated highly over and above conditions, $r(182) = 0.76$, $p < .001$.

6.3. Discussion

The results of Study 2 extend the findings of Study 1 concerning the role of confronter gender and appraisal of the confrontation. Regarding Hypothesis 1, the text of the male confronter was not evaluated significantly more positively than the text of the female confronter, indicating only a small effect on overall privilege attitudes. However, we found that men justified male privilege less when the author was male and that male confronters triggered less backlash responses. In line with Hypothesis 2 and results of Study 1, the male confronter was also evaluated more positively.

In the close coding of text-based thought protocols we found four main categories of privilege responses: Accepting male privilege and calling for social change, justifying male privilege and distancing from responsibility, rejecting male privilege and gender inequality, as well as backlash. The first three categories can be roughly equated to the three d's in the 3D model: Acceptance is similar to the dismantle response, justification is close to distancing, and rejection has most overlaps with denial. However, our codes are somewhat broader. For example, acceptance also includes acceptance without a pledge for social change and justification also includes legitimization of inequality. Additionally, backlash has not been considered in the 3D model but was an important outcome of male privilege confrontations here. Therefore, we examined backlash further in Study 3, where we also devised new bipolar privilege response scales based on the acceptance, justification, and rejection response categories.

The thought protocols did not provide direct information on the extent of image threat and we could not differentiate between image and moral appraisals. We address these limitations in Study 3.

7. Study 3

The appraisal-based approach is rooted in the idea that confrontations with ingroup privilege can be perceived as threatening the ingroup's image (image appraisal) and/or the ingroup's moral integrity (moral appraisal). These different appraisals of the confrontation should then be related to different dimensions of shame (i.e., image shame and moral shame). Finally, image shame should lead to negative privilege attitudes and moral shame to constructive privilege attitudes. Therefore, we predicted that the image appraisal path leads to negative, and the moral appraisal path will lead to positive reactions toward male privilege (Hypothesis 1). We also expected that the moral appraisal path explains more positive privilege attitudes in the male (compared to the female) confronter condition (Hypothesis 2) and that the male confronter is evaluated more positively than the female confronter, explained through the image appraisal path (Hypothesis 3). Study 3 also provides a second test of the 3D model approach with meritocratic threat and group-image threat as mediators and a test of guilt as an alternative emotional path (reported in the supplementary materials section C). These hypotheses and analyses were preregistered. To

measure privilege attitudes more in line with the relevant concerns that men have while thinking about male privilege, we designed items to tap into the different categories of reactions we found in Study 2: acceptance or feminism and privilege, justification of status, and rejection of responsibility. These three categories of reactions to the confrontation also relate to the three outcomes proposed in the 3D model, although they are more specific: Acceptance of feminism and privilege could be interpreted as an inverse measure of denial of privilege, justification of personal status as a direct measure of distancing from privilege, and rejection of responsibility as an inversed measure of dismantling.

7.1. Method

7.1.1. Pretest

An initial set of $N = 49$ male participants was collected via Prolific Academic to evaluate the performance of the new scales used in this Study. Results and subsequent changes to the items are reported in the supplementary material (section C). The pretest sample was not used for the main analyses reported below.

7.1.2. Participants

We preregistered to collect $N = 700$ participants to retain a sample size equivalent to that of Study 1. We collected data via an online questionnaire circulated to adult male German residents through Prolific Academic ($n = 453$ completed surveys) and LoopsterPanel ($n = 271$ completed surveys) for a total of $N = 724$ completed surveys. Participants from LoopsterPanel were added because Prolific Academic's pool of German males was not sufficiently large and data collection slowed down significantly after the first days. Data from $n = 33$ participants who reported that they had not been assigned male sex at birth, hence not categorizing as "cis-male", were excluded. Another $n = 56$ were excluded, because participants failed an attention check similar to Study 1, and $n = 10$ were excluded based on a relative speed index above 2 (Leiner, 2019b). The final sample consisted of $N = 623$ cis-men with a mean age of 37.43 ($SD = 15.09$) years and above average subjective social status on a scale from 1 to 10 ($M = 5.99$, $SD = 1.63$).

7.1.3. Materials & procedure

Factor analyses were conducted for all dependent variables (threat, shame, privilege awareness and privilege attitudes polarities) using principal axis and parallel analysis, confirming the expected structures of these scales (see supplementary materials section C). Correlations between the main dependent variables appear in Table 4.

Demographics. Participants provided demographic data on their age, gender, migration background, nationality, and subjective social status (Hoebel et al., 2015) at the beginning of the survey.

Manipulation of confrontation and gender. We used the same texts as in Study 1 and 2 for the three conditions male confronter, female confronter, and information. The confronters' names were taken from Study 2.

Threat. Group-image threat was measured similarly to Study 1 (ρ = 0.73). Only the beginning of the items "While reading the text, I had the impression…" was moved to the instruction for brevity. Moral threat was measured with two additional items, "… that men have a moral deficit" and "… that men are morally failing" (ρ = 0.88). We adapted items of the meritocratic threat scale from Study 1 to tap more specifically into text-based reactions instead of a general political attitude: "I feel threatened by the statement that my successes are due to my gender" and "I feel threatened by the statement that my social status is based on my gender" (ρ = 0.95). These items were measured on a 7-point scale from 1 (strongly disagree) to 7 (strongly agree).

Shame and additional emotional appraisals. Image shame and moral shame were measured with two items each based on the scales used by Allpress et al. (2014). Items measuring image shame were, "I am ashamed because men are portrayed negatively" and "I am ashamed of the negative impression men give" (ρ = 0.86). Items measuring moral shame were, "I am ashamed of the undeserved advantages men have" and "I feel ashamed because of the unjustified preferential treatment of men" (ρ = 0.92). We used one item to measure guilt, "I feel guilty when thinking about the text". Guilt was used to test whether image and moral shame effects are robust for the inclusion of guilt (see Allpress et al., 2014). We additionally measured anger with one item, "I am angry about the way the text portrays men". Anger and guilt were not central to the current manuscript (see supplementary materials section C for results on these emotions). All emotions were measured on a 9-point scale from 1 (not at all) to 9 (very much).

Author evaluation. Author evaluation was measured as in Study 1.

Privilege attitudes. Text evaluation was measured as in Study 2. Male privilege awareness (α = 0.81) was measured as in Study 1. Additionally, we created six bi-polar items to tap into different reactions toward male privilege based on the response categories found in Study 2. Two items measured backlash against/acceptance of feminism (e.g., "Feminism goes too far/Feminism does not go far enough"; ρ = 0.79). Two items measured justification/acceptance of personal status (e.g., "I have mainly disadvantages/advantages in my life because I am a man"; ρ = 0.66). Two items measured rejection/acceptance of responsibility to dismantle male privilege (e.g., "I don't want/want to take responsibility for dismantling male privilege; ρ = 0.78). These items were measured on a scale from −3 to 3 and were then recoded to 1 to 7. For easier interpretation, justification of personal status and rejection of responsibility responses were reversed so that high values represent more justification (of personal status) and stronger rejection (of individual responsibility).

Attention check and additional variables. Political orientation was measured on a scale from 0 (left) to 100 (right) near the end of the survey to check the comparability of the conditions (see supplementary materials section C). Finally, as attention check, we asked participants to select one out of three statements that best described the text they had read in the beginning.

The correlation between author and text evaluation was extremely high, $r(621) = 0.86$, $p < .001$. However, we did not aggregate them to retain consistency between studies regarding this outcome. We focus on author evaluation in the results. An extended table with text evaluation

Table 4
Correlation matrix with relevant dependent variables.

		1	2	3	4	5	7	8	9	10
1	Group-Image Threat	4.33 (1.57)								
2	Moral Threat	0.40***	3.46 (1.67)							
3	Meritocratic Threat	0.29***	0.15***	3.41 (1.93)						
4	Image Shame	0.21***	0.32***	0.33***	3.49 (2.23)					
5	Moral Shame	0.08*	0.33***	0.08	0.63***	3.67 (2.30)				
7	Privilege Awareness	−0.14***	0.10*	−0.32***	0.08***	0.42***	4.42 (1.55)			
8	Acceptance	−0.10*	0.18***	−0.32***	0.17***	0.49***	0.69***	3.61 (1.51)		
9	Justification	0.12**	−0.06	0.16***	−0.09*	−0.28***	−0.51***	−0.49***	3.87 (1.25)	
10	Rejection	0.04	−0.19***	0.16***	−0.25***	−0.53***	−0.54***	−0.66***	0.48***	3.81 (1.55)
11	Author Evaluation	−0.24***	0.06	−0.26***	0.06	0.32***	0.47***	0.55***	−0.35***	−0.49***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Means (and standard deviation) appear on the diagonal.

and additional variables appears in the supplementary materials (section C).

7.2. Results

To test the total effects stated in Hypotheses 1–3, we conducted univariate ANOVAs of condition on privilege attitudes and author evaluation (see Table 5 for results across all relevant dependent variables). Condition significantly affected privilege awareness but only the effect between the female confronter and the information condition was significant, $t(620) = -2.52$, $p_{tukey} = 0.03$, $d = -0.24$ (95% CI [−0.43–0.05]). As such, the total effects predicted in Hypothesis 1 and 2 were only partially confirmed. Consistent with Hypothesis 3 (and Study 1 and 2), there was a significant effect on author evaluation. Post-hoc tests revealed that the male confronter was evaluated more positively than the female confronter, $t(620) = -5.18$, $p_{tukey} < 0.001$, $d = -0.52$ (95% CI [−0.72–0.32]), and the work group in the information condition was evaluated more positively than the female author, $t(620) = -7.68$, $p_{tukey} < 0.001$, $d = -0.74$ (95% CI [−0.93–0.55]). The difference in author evaluations between the male confronter and the information condition was not significant, $t(620) = -2.23$, $p_{tukey} = 0.07$, $d = -0.22$ (95% CI [−0.42–0.03]). There was an effect of condition on text evaluation, $F(2,620) = 20.28$, $p < .001$, $\eta_p^2 = 0.06$. Similar to author evaluation, we found that the confrontation was evaluated more negatively when it came from a female versus male author, $t(620) = -4.26$, $p_{tukey} < 0.001$, $d = -0.42$ (95% CI [−0.62–0.23]). The text of the female confronter was also more negatively evaluated than the information, $t(620) = -6.22$, $p_{tukey} < 0.001$, $d = -0.60$ (95% CI [−0.79–0.41]). In contrast, the evaluation of the confrontation written by the male author was not significantly lower than the evaluation of the information, $t(620) = -1.74$, $p_{tukey} = 0.19$, $d = -0.17$ (95% CI [−0.37 0.02]).

Our manipulation affected the levels of group-image and moral threat as expected and, consistent with Study 1, there was no effect on meritocratic threat. Results of the 3D model configuration (with meritocratic and group-image threat as parallel mediators) also mirror those of Study 1 and are reported in the supplementary material (section C). Regarding emotional appraisals, we found no total effect on both dimensions of shame.

To test the effect of the different appraisals of the confrontation, a path model was computed using the pathj module (Gallucci, 2021) for jamovi (The jamovi project, 2021) with maximum likelihood estimation and robust standard errors. Group-image threat and moral threat were regressed on the exogenous factor condition, image shame was regressed on group-image threat (image appraisal path) and moral shame was regressed on moral threat (moral appraisal path). Privilege attitudes (awareness, acceptance, justification, rejection) and author evaluation were regressed on both dimensions of shame. As in Study 1, author evaluation was also regressed directly on the condition variable, all continuous variables were mean-centered, and the condition variable was difference-coded to represent the two relevant contrasts (information-confrontations and male-female confronter). Furthermore, we allowed covariance of the residuals of both dimensions of threat and both dimensions of shame. The path model had good fit, $\chi^2(24) = 94.15$, $p < .001$, CFI = 0.97, RMSEA = 0.07 ($p = .02$, 95% CI [0.05 0.08]), SRMR = 0.05. Fig. 3 represents the results graphically. Table 6 provides an overview of all indirect effects (IE). Details for all paths can be found in the supplementary material (section C).

Consistent with Hypothesis 1, the confrontations were perceived as more threatening than the information and the image appraisal path explained more negative privilege attitudes and author evaluations (see indirect effects in Table 6: IE11-IE15 and IE21–24), whereas the moral appraisal path explained more positive privilege attitudes and author evaluations (IE16–20 and IE26–29). Deviating from Hypothesis 2, the female confronter triggered both more group-image threat, $b = -0.64$, 95% CI [−0.92–0.35], $\beta = -0.16$, $p < .001$, and more moral threat, $b = -0.39$, 95% CI [−0.72–0.06], $\beta = -0.09$, $p = .02$. Indirect effects of this contrast on privilege attitudes and author evaluation were significant through the negative image and the positive moral appraisal path (IE1–10). In line with Hypothesis 3, the indirect effect on author evaluation through the image appraisal path was significant (IE25). We also found an indirect effect on author evaluation through the moral appraisal path (IE30).

Table 5
Descriptive information on relevant dependent variables and summary of ANOVA's.

	Female (N = 211)	Male (N = 190)	Information (N = 222)	ANOVA		
	M (SD)	M (SD)	M (SD)	F (2,620)	p	η_p^2
Meritocratic Threat	3.60 (2.02)	3.45 (1.92)	3.24 (1.86)	1.83	0.16	0.006
Group-image Threat	4.87 (1.37)	4.23 (1.55)	3.90 (1.62)	22.79	< 0.001	0.07
Moral Threat	3.78 (1.71)	3.39 (1.64)	3.21 (1.63)	6.71	0.001	0.02
Image Shame	3.55 (2.22)	3.45 (2.19)	3.48 (2.27)	0.10	0.91	0.00
Moral Shame	3.60 (2.23)	3.60 (2.23)	3.79 (2.43)	0.53	0.59	0.002
Awareness	4.21 (1.50)	4.48 (1.65)	4.58 (1.50)	3.34	0.04	0.01
Acceptance	3.43 (1.42)	3.64 (1.57)	3.75 (1.54)	2.54	0.08	0.008
Justification	4.23 (1.19)	4.07 (1.39)	4.08 (1.19)	1.08	0.34	0.003
Rejection	4.35 (1.45)	4.23 (1.60)	3.99 (1.58)	3.02	0.05	0.01
Auhtor Evaluation	30.85 (24.18)	44.02 (27.52)	49.63 (24.73)	30.78	< 0.001	0.09

7.3. Discussion

The results of Study 3 largely confirmed our hypotheses with some additional effects: First, although the differences in privilege attitudes between the confrontations and information conditions were not large, we found evidence that confrontations trigger more image threat, which were related to image shame and negative privilege attitudes (Hypothesis 1). This finding confirmed the results of Study 1. We also found evidence for the additional moral appraisal path through moral threat and moral shame, which was related to more constructive privilege attitudes. Second, we hypothesized that a male confronter would trigger more constructive privilege attitudes than a female confronter, which would be explained via the moral appraisal path (Hypothesis 2). However, we only found a small effect of our manipulation on privilege awareness, which was due to the mean difference in the female confronter and information conditions. Path analysis revealed that the female confronter was perceived as more threatening on both the group-image *and* moral dimension. As these paths had opposing implications for privilege attitudes, this might explain the small total effects of our manipulation. Third, the results were consistent with our hypothesis that the male confronter would be evaluated more positively than the female confronter, which could be explained through the image appraisal path (Hypothesis 3). We also found that the moral appraisal path was contributing positively to author evaluation. Fourth, confirming the results of Study 1, meritocratic threat was not affected by the confrontation but was negatively related to privilege attitudes. Taken together, Study 3 highlights the relevance of distinguishing between appraisals of ingroup privilege as rooted in concerns for the ingroup's morality and for the ingroup's image.

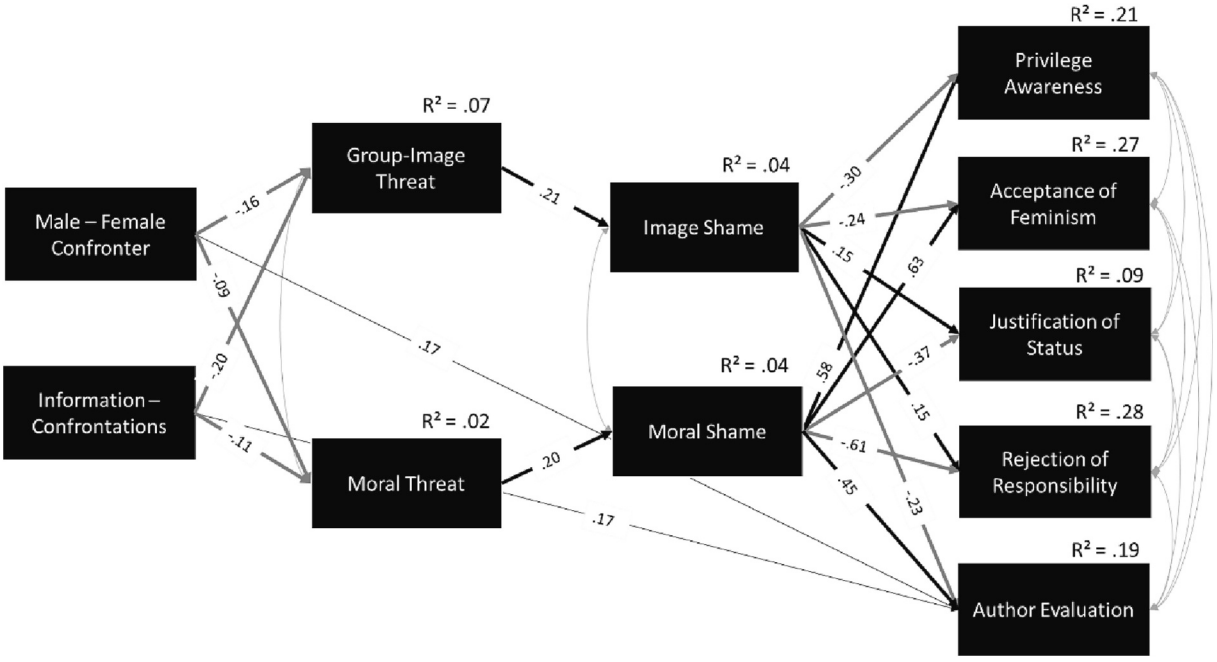


Fig. 3. Path model of the appraisal-based approach.
Note. Estimates are standardized. All paths are significant $p < .05$. Grey arrows represent negative effects. Narrow arrows represent the estimated direct effects between condition and author evaluation. Curved arrows represent estimated residual covariance between the mediators.

Table 6
Overview of indirect effects.

	Path	Outcome	*b*	LCI	UCI	β	*p*
Male-Female confronter contrast							
IE1	Image	Awareness	0.04	0.01	0.06	0.01	0.003
IE2	Image	Acceptance	0.03	0.01	0.05	0.01	0.002
IE3	Image	Justification	−0.02	−0.03	0.00	−0.01	0.018
IE4	Image	Rejection	−0.02	−0.04	0.00	−0.01	0.012
IE5	Image	Evaluation	0.52	0.17	0.86	0.01	0.003
IE6	Moral	Awareness	−0.04	−0.08	0.00	−0.01	0.033
IE7	Moral	Acceptance	−0.04	−0.09	0.00	−0.01	0.033
IE8	Moral	Justification	0.02	0.00	0.04	0.01	0.035
IE9	Moral	Rejection	0.04	0.00	0.08	0.01	0.033
IE10	Moral	Evaluation	−0.55	−1.07	−0.04	−0.01	0.036
Information-Confrontations contrast							
IE11	Image	Awareness	0.04	0.02	0.07	0.01	0.002
IE12	Image	Acceptance	0.03	0.01	0.05	0.01	0.001
IE13	Image	Justification	−0.02	−0.03	0.00	−0.01	0.021
IE14	Image	Rejection	−0.02	−0.04	0.00	−0.01	0.013
IE15	Image	Evaluation	0.53	0.18	0.88	0.01	0.003
IE16	Moral	Awareness	−0.04	−0.07	−0.01	−0.01	0.009
IE17	Moral	Acceptance	−0.04	−0.08	−0.01	−0.01	0.009
IE18	Moral	Justification	0.02	0.01	0.04	0.01	0.012
IE19	Moral	Rejection	0.04	0.01	0.08	0.01	0.010
IE20	Moral	Evaluation	−0.54	−0.95	−0.13	−0.01	0.011
Indirect effects of group-image and moral threat over and above condition							
IE21	Image	Awareness	−0.06	−0.09	−0.03	−0.06	<0.001
IE22	Image	Acceptance	−0.05	−0.07	−0.03	−0.05	<0.001
IE23	Image	Justification	0.03	0.01	0.04	0.03	0.007
IE24	Image	Rejection	0.03	0.01	0.05	0.03	0.003
IE25	Image	Evaluation	−0.81	−1.24	−0.39	−0.05	<0.001
IE26	Moral	Awareness	0.11	0.07	0.15	0.12	<0.001
IE27	Moral	Acceptance	0.11	0.07	0.16	0.13	<0.001
IE28	Moral	Justification	−0.06	−0.08	−0.03	−0.07	<0.001
IE29	Moral	Rejection	−0.11	−0.15	−0.07	−0.12	<0.001
IE30	Moral	Evaluation	1.41	0.86	1.96	0.09	<0.001

Note. IE1 to IE30 = labels of indirect effects.

8. General discussion

Men are better allies in the quest for dismantling sexist structures when they are aware of systemic ingroup privilege and acknowledge that there are structures that invisibly and automatically provide men with benefits that women do not enjoy (Droogendyk et al., 2016;

Kutlaca, Radke, et al., 2020; Radke, Kutlaca, Siem, Wright, & Becker, 2020). Thus, exploring what makes male privilege confrontations more (or less) effective is an important avenue of research. We examined different dimensions of identity threat that have been proposed by privilege confrontation research (i.e., to meritocratic beliefs, to the ingroup's image) and by research on ingroup transgressions and moral failure (i.e., to ingroup morality). We also investigated whether the effects of a confrontation with male privilege are shaped by the ingroup (male) versus outgroup (female) status of the confronter.

Results of Studies 1 and 3 confirmed that being confronted with ingroup privilege is associated with perceiving group-image threat. However, contrary to the 3D model's proposition (Knowles et al., 2014), group-image threat was *negatively* associated with attitudes toward male privilege and toward the confronters. Although we found that meritocratic threat was negatively related to attitudes toward male privilege, we did not find evidence that this threat was affected by the confrontation. Instead, Study 3 demonstrated that moral threat - > moral shame (the moral appraisal path) was related to constructive attitudes toward male privilege and authors whereas group-image threat - > image shame (the image appraisal path) had negative effects.

In all three studies the female confronter was evaluated more negatively than the male confronter. Study 3 specified that only the evaluation of the female but not of the male confronter was more negative than the evaluation of the work group behind the information. This is strong evidence of a specific bias against female confronters of male privilege and is in line with previous literature on bias against female sexism confronters (e.g., Czopp & Monteith, 2003; Glick, 2014). While text evaluations were strongly related to author evaluations, our more specific privilege attitudes measures were largely unaffected by our experimental manipulation. However, after confrontation by a male (versus female) author, men were less likely to attempt to justify their male privilege and engaged in less backlash (Study 2), and were less prone to reject responsibility (Study 3).

In Study 3, the female confronter unexpectedly triggered both the positive (moral) and the negative (image) appraisal routes more strongly than the male confronter, and we found the same pattern in the information versus confrontation contrast. As the moral and image appraisal path have opposite effects on privilege attitudes, this could explain the null-effects on these outcomes. Taken together, our results indicate that image concerns undermine the effectiveness of privilege interventions. While we could demonstrate the importance of distinguishing between group-image and moral threat, results are somewhat limited regarding the question how these threats can be triggered separately. Nevertheless, we propose possible strategies in the section on implications for male privilege interventions below. Table 7 provides an overview of how the results in Study 1–3 relate to the predictions of the appraisal-based model introduced above.

Table 7
Summary of results in relation to the appraisal-based model.

	Evidence supporting the appraisal-based model	Evidence conflicting with the model
Study 1	The male confronter triggered less group-image threat and group-image threat was negatively related to privilege attitudes. Meritocratic threat was not affected, indicating additional mechanisms not considered in the 3D model. Privilege remorse was positively associated with privilege awareness, even when controlling for group-image threat.	Indirect effects through group-image threat only approached significance and total effects on privilege attitudes were non-significant.
Study 2	Broad range of responses to the confrontations beyond mere meritocratic responses (i.e., claiming life hardships) Less justification and backlash in the male confronter condition.	Evaluation of the male confronter's text were less favorable, but the difference was not significant.
Study 3	Moral appraisal path was associated with positive privilege attitudes and image appraisal path was associated with negative privilege attitudes.	There were again no total effects on privilege attitudes (except for justification of privilege). Female confronter triggered not only more image but also more moral threat.

8.1. Theoretical implications

The positive role of the moral and the negative role of the image appraisal path are in line with the findings that a desire to be perceived positively by an outgroup leads to less constructive outcomes than an ingroup-focused motivation, such as adhering to moral norms (Aquino & Reed, 2002; Gausel & Leach, 2011). It also aligns with research on moral failure, where moral threat and shame were associated with the willingness to apologize for ingroup transgressions, to support reparations, and with more moral engagement (Cibich et al., 2016; Wenzel, Woodyatt, & McLean, 2020). Integrating research on privilege interventions with the theoretical approach of research on ingroup transgressions and moral failure therefore proved to be a promising way forward for the field.

The clearer examination of the role of moral and image appraisal of privilege confrontations also allows us to shed new light on the equivocal findings of previous research regarding the responses to ingroup versus outgroup confronters of sexism. Although in some experiments men derogated male confronters more strongly than female confronters (e.g., Kutlaca et al., 2020), other research found that men were evaluated more positively and were perceived as more legitimate (e.g., Drury & Kaiser, 2014). One possible reason for these contradicting findings could be the presence of hidden moderators that exacerbated the salience of moral and/or image concerns: In the experiments by Kutlaca et al. (2020), for example, the confrontation took place in public spaces (e.g., drugstore, paintball club) and participants were in the role of a bystander of the confrontation (instead of being addressed themselves). This may have created a greater sensitivity to image concerns than our direct confrontation through a newspaper article did (see also Monin, Sawyer, & Marquez, 2008). Therefore, ingroup/outgroup positionality of the confronter might play a less specific role for the appraisal of a private (versus public) confrontation. Future research is needed to investigate this potential interaction between confrontation context and confronter gender on moral and image appraisal.

Beyond the presence of hidden moderators there is another possible explanation for the finding that, contrary to our original predictions, the female confronter increased the relevance of image *and* moral concerns. Previous research found that when status differences are delegitimized, members of advantaged groups seek confirmation of their social acceptance (i.e., their morality) from disadvantaged groups (e.g., Siem, Oettingen, Mummendey, & Nadler, 2013). A female confronter of male privilege may violate this need, both leading men to reflect on the morality of male privilege but also triggering a need to save face. Whether image or moral concerns take precedence might depend on the impact of contextual (e.g., presence of bystanders or immediate ability to redeem oneself), inter-individual (e.g., relationship closeness to the confronter), and individual factors (e.g., political attitudes). Thus, our original predictions and the appraisal-based model would need to be extended by these moderators to predict moral versus image appraisal.

The stronger derogation of the female compared to the male confronter is in line with findings on backlash against female comforters of sexism (Czopp & Monteith, 2003; Rasinski & Czopp, 2010). While there were some effects regarding measures close to the responses defined by the 3D model, such as justification (i.e., distancing) in Study 2 or rejection of feminism (i.e., inverse of dismantling) in Study 3, these results were somewhat inconsistent. Recent research proposed an extension of the 3D model (4D model) by adding defensiveness (Shuman, van Zomeren, Saguy, & Knowles, 2022). Building on this, research might consider investigating derogation as a separate, potentially more

primary response to privilege confrontations.

8.2. Implications for male privilege interventions

By clearly separating the moral and image dimensions from the 3D model's definition of group-image threat, we found that image *and* moral threat was higher in the female (versus male) and in the confrontations (versus information) contrasts. This "double threat" to image and morality can explain the lack of total effects on attitudes toward privilege. For practical interventions, this means that confrontations with male privilege might be more effective when they trigger moral concerns but avoid threatening men's group image. We had originally expected that the male confronter was especially well positioned to do so. However, our findings are more complicated than that: Although male confronters seem to be protected from derogation by virtue of their ingroup status, they are not necessarily more effective.

Feminist theory emphasizes that gender inequality and rigid gender roles are harmful to both women *and* men (e.g., Hooks, 2005). It might be fruitful for interventions to focus on intersecting systems of advantage and disadvantage and how they are detrimental even to those who seem to profit at first glance (Coston & Kimmel, 2012; Ehrke, Ashoee, Steffens, & Louvet, 2020), in order to specifically trigger moral but not image related threat. For example, interventions may emphasize how stereotypical views of men as more competent but less emotional have negative consequences for them (Flynn, Hollenstein, & Mackey, 2010) or how men's reduced willingness to seek help can explain higher rates of suicide among men (Oliffe et al., 2020). The lower inclination to derogate fellow men might aid male confronters to relay the moral implications of ingroup privilege through such a more system-critical perspective. Another important avenue for future research concerns the experience of structural (dis-)advantage of non-cis men, which we did not investigate in the present paper (for more on such inclusive research strategies see Chetkovich, 2019).

8.3. Limitations

The strength of the current paper lies in the fact that we used various measures and methods across the three studies to triangulate our hypotheses. This decreases the likelihood of alternative explanations for our findings, contributing to their internal validity. Still, some limitations, especially concerning the validity of the measurement of attitudes toward male privilege should be addressed. The operationalization of privilege awareness and privilege remorse in Study 1 were adapted from a validated White Privilege Attitude scale (Pinterits et al., 2009). Using unidimensional scales to measure privilege attitudes, however, conflates knowledge about group-based advantages with the willingness to admit them. Additionally, social desirability may influence responses directly after a privilege intervention, especially as participants are likely to understand the researcher's interest in the intervention condition. This is a considerable problem for most experimental research in this field. For this reason, we used a less direct measure of privilege attitude in Study 2 (evaluation of the text) as well as polarity profiles in Study 3, where negative and positive reactions to privilege can be expressed equally well. The consistency of results across these different operationalizations provides confidence that response biases are not (solely) responsible for the effects.

The differences between male and female confronters were not large when it comes to the most important outcomes: making men aware of male privileges and their politicization as allies (Kutlaca, Radke, et al., 2020). However, our focus on evaluating different mechanisms revealed that there are positive and negative dimensions of identity threat. This has potential for future research, which should test whether male privilege interventions designed to specifically trigger appraisals of moral threat can inspire political solidarity among men. Another important question concerns the longevity of intervention effects. Practical privilege interventions do seem to have effects on attitudes toward privilege and the disadvantaged outgroup a few weeks later (Ehrke et al., 2020). However, we lack crucial information about the factors that contribute to the long-term effectiveness of privilege interventions. Intervention tournaments are a method where multiple interventions that systematically differ (e.g., in how they appeal to the ingroup's morality) are tested against a single control (Hameiri & Moore-Berg, 2022). Using intervention tournaments in conjunction with a longitudinal design would help distinguish between components of privilege interventions that contribute to negative versus positive outcomes and to short and/or long-term effects. Longitudinal designs would also address the problem of assumed but untested causality between the mediators and outcomes in the present research.

9. Conclusion

The current paper demonstrates that the identity threat following privilege confrontations can be separated into concerns about the ingroup's image and the ingroup's morality. We found that image concerns have negative and moral concerns have positive implications for subsequent privilege attitudes. Our findings also suggest that male allies trigger less backlash and less group-image threat. Male allies might therefore harness their ingroup position to inspire moral concerns about male privilege among their peers with less potential to receive backlash. However, the difference between male and female confronters and between confrontation and information on privilege attitudes were mostly non-significant. This might be explained by the finding that confrontations (versus information) and female confronters (versus male confronters) trigger more group-image *and* more moral threat. The three experiments reported here therefore indicate a need for future research to investigate which components of male privilege interventions trigger appraisal of threatened ingroup morality while keeping in check concerns about the ingroup's image.

Open practice

The pre-registrations, supplementary material, manipulation material, data sets and result outputs are stored in an Open Science Framework repository: https://osf.io/pysdw/?view_only=69faa4cdedac49bd97baafa6d1eda76b

Declaration of Competing Interest

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

I have shared a link to the data and supplementary material in the Open Practice section in the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2023.104492.

References

Allpress, J. A., Brown, R., Giner-Sorolla, R., Deonna, J. A., & Teroni, F. (2014). Two faces of group-based shame: Moral shame and image shame differentially predict positive and negative orientations to ingroup wrongdoing. *Personality and Social Psychology Bulletin, 40*(10), 1270–1284. https://doi.org/10.1177/0146167214540724

Ancis, J. R., & Szymanski, D. M. (2001). Awareness of white privilege among white counseling trainees. *The Counseling Psychologist, 29*(4), 548–569.

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423–1440. https://doi.org/10.1037//0022-3514.83.6.1423

Barajas, S. (2022). Unearned advantages? Redefining privilege in light of childhood. *Children's Geographies, 20*(1), 24–36. https://doi.org/10.1080/14733285.2021.1912295

Brambilla, M., Sacchi, S., Pagliaro, S., & Ellemers, N. (2013). Morality and intergroup relations: Threats to safety and group image predict the desire to interact with outgroup and ingroup members. *Journal of Experimental Social Psychology, 49*(5), 811–821. https://doi.org/10.1016/j.jesp.2013.04.005

Branscombe, N. R. (1998). Thinking about one's gender group's privileges or disadvantages: Consequences for well-being in women and men. *British Journal of Social Psychology, 37*, 167–184. https://doi.org/10.1111/j.2044-8309.1998.tb01163.x

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: Context, commitment, content* (pp. 35–58). Blackwell Science.

Case, K. A., Hensley, R., & Anderson, A. (2014). Reflecting on heterosexual and male privilege: Interventions to raise awareness. *Journal of Social Issues, 70*(4), 722–740. https://doi.org/10.1111/josi.12088

Chetkovich, C. (2019). How non-binary gender definitions confound (already complex) thinking about gender and public policy. *Journal of Public Affairs Education, 25*(2), 226–252. https://doi.org/10.1080/15236803.2018.1565050

Cibich, M., Woodyatt, L., & Wenzel, M. (2016). Moving beyond "shame is bad": How a functional emotion can become problematic. *Social and Personality Psychology Compass, 10*(9), 471–483. https://doi.org/10.1111/spc3.12263

Coston, B. M., & Kimmel, M. (2012). Seeing privilege where it isn't: Marginalized masculinities and the intersectionality of privilege. *Journal of Social Issues, 68*(1), 97–111.

Coston, B. M., & Kimmel, M. (2013). White men as the new victims: Reverse discrimination cases and the Men's rights movement. *Nevada Law Journal, 13*(2), 368–385.

Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin, 29*(4), 532–544. https://doi.org/10.1177/0146167202250923

Dörr, J. (2019, March 7). Ja, alle Männer. https://www.sueddeutsche.de/kultur/mansplaining-feminismus-maennlichkeit-1.4357864.

Droogendyk, L., Wright, S. C., Lubensky, M., & Louis, W. R. (2016). Acting in solidarity: Cross-group contact between disadvantaged group members and advantaged group allies. *Journal of Social Issues, 72*(2), 315–334. https://doi.org/10.1111/josi.12168

Drury, B. J., & Kaiser, C. R. (2014). Allies against sexism: The role of men in confronting sexism. *Journal of Social Issues, 70*(4), 637–652. https://doi.org/10.1111/josi.12083

Ehrke, F., Ashoee, A., Steffens, M. C., & Louvet, E. (2020). A brief diversity training: Raising awareness of ingroup privilege to improve attitudes towards disadvantaged outgroups. *International Journal of Psychology, 55*(5), 732–742. https://doi.org/10.1002/ijop.12665

Eliezer, D., & Major, B. (2012). It's not your fault: The social costs of claiming discrimination on behalf of someone else. *Group Processes & Intergroup Relations, 15*(4), 487–502. https://doi.org/10.1177/1368430211432894

Ellemers, N., Pagliaro, S., & Barreto, M. (2013). Morality and behavioural regulation in groups: A social identity approach. *European Review of Social Psychology, 24*(1), 160–193. https://doi.org/10.1080/10463283.2013.841490

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology, 53*, 161–186.

Ellemers, N., & van den Bos, K. (2012). Morality in groups: On the social-regulatory functions of right and wrong. *Social and Personality Psychology Compass, 6*(12), 878–889. https://doi.org/10.1111/spc3.12001

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Flood, M., Dragiewicz, M., & Pease, B. (2021). Resistance and backlash to gender equality. *Australian Journal of Social Issues, 56*(3), 393–408. https://doi.org/10.1002/ajs4.137

Flynn, J. J., Hollenstein, T., & Mackey, A. (2010). The effect of suppressing and not accepting emotions on depressive symptoms: Is suppression different for men and women? *Personality and Individual Differences, 49*(6), 582–586. https://doi.org/10.1016/j.paid.2010.05.022

Fortgang, T. (2014, May 2). Why I'll never apologize for my white male privilege. *Time*. https://time.com/85933/why-ill-never-apologize-for-my-white-male-privilege/.

Galea, N., & Gaweda, B. (2018). (De)constructing the masculine blueprint: The institutional and discursive consequences of male political dominance. *Politics & Gender, 14*(2), 276–282. https://doi.org/10.1017/S1743923X18000168

Gallucci, M. (2021). PATHj: Jamovi path analysis [computer software]. https://pathj.github.io.

Gausel, N., & Leach, C. W. (2011). Concern for self-image and social image in the management of moral failure: Rethinking shame. *European Journal of Social Psychology, 41*(4), 468–478. https://doi.org/10.1002/ejsp.803

Gausel, N., Leach, C. W., Vignoles, V. L., & Brown, R. (2012). Defend or repair? Explaining responses to in-group moral failure by disentangling feelings of shame, rejection, and inferiority. *Journal of Personality and Social Psychology, 102*(5), 941–960. https://doi.org/10.1037/a0027233

Gausel, N., Vignoles, V. L., & Leach, C. W. (2016). Resolving the paradox of shame: Differentiating among specific appraisal-feeling combinations explains pro-social and self-defensive motivation. *Motivation and Emotion, 40*(1), 118–139. https://doi.org/10.1007/s11031-015-9513-y

Glick, P. (2014). Commentary: Encouraging confrontation. *Journal of Social Issues, 70*(4), 779–791. https://doi.org/10.1111/josi.12091

Hameiri, B., & Moore-Berg, S. L. (2022). Intervention tournaments: An overview of concept, design, and implementation. *Perspectives on Psychological Science*. https://doi.org/10.1177/17456916211058090, 17456916211058090.

Hoebel, J., Müters, S., Kuntz, B., Lange, C., & Lampert, T. (2015). Messung des subjektiven sozialen status in der Gesundheitsforschung mit einer deutschen version der MacArthur scale [measuring subjective social status in health research with a German version of the MacArthur scale]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz, 58*(7), 749–757. https://doi.org/10.1007/s00103-015-2166-x

Hooks, B. (2005). *The will to change: Men, masculinity, and love*. Washington Square Press.

Jones, L., Chałupnik, M., Mackenzie, J., & Mullany, L. (2022). 'STFU and start listening to how scared we are': Resisting misogyny on twitter via #NotAllMen. *Discourse, Context & Media, 47*, Article 100596. https://doi.org/10.1016/j.dcm.2022.100596

Kaptein, M., & van Helvoort, M. (2019). A model of neutralization techniques. *Deviant Behavior, 40*(10), 1260–1285. https://doi.org/10.1080/01639625.2018.1491696

Knowles, E. D., Lowery, B. S., Chow, R. M., & Unzueta, M. M. (2014). Deny, distance, or dismantle? How white Americans manage a privileged identity. *Perspectives on Psychological Science, 9*(6), 594–609. https://doi.org/10.1177/1745691614554658

Kutlaca, M., Becker, J., & Radke, H. R. M. (2020). A hero for the outgroup, a black sheep for the ingroup: Societal perceptions of those who confront discrimination. *Journal of Experimental Social Psychology, 88*, Article 103832. https://doi.org/10.1016/j.jesp.2019.103832

Kutlaca, M., Radke, H. R. M., Iyer, A., & Becker, J. C. (2020). Understanding allies' participation in social change: A multiple perspectives approach. *European Journal of Social Psychology, 50*(6), 1248–1258. https://doi.org/10.1002/ejsp.2720

Leach, C. W., Bilali, R., & Pagliaro, S. (2015). Groups and morality. In M. Mikulincer, P. R. Shaver, J. F. Dovidio, & J. A. Simpson (Eds.), *APA handbook of personality and social psychology, volume 2: Group processes* (pp. 123–149). American Psychological Association. https://doi.org/10.1037/14342-005.

Leiner, D. J. (2019a). Sosci survey (version 3.1.06) [computer software]. https://www.soscisurvey.de.

Leiner, D. J. (2019b). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in Internet surveys. *Survey Research Methods, 13*(3), 229–248. https://doi.org/10.18148/SRM/2019.V13I3.7403

Maricourt, C., & Burrell, S. R. (2022). #MeToo or #MenToo? Expressions of backlash and masculinity politics in the #MeToo era. *Journal of Men's Studies, 30*(1), 49–69. https://doi.org/10.1177/10608265211035794

Marques, J. M., Abrams, D., Paez, D., & Martinez-Taboada, C. (1998). The role of categorization and in-group norms in judgments of groups and their members. *Journal of Personality and Social Psychology, 75*(4), 976–988. https://doi.org/10.1037/0022-3514.75.4.976

Marques, J. M., & Paez, D. (1994). The 'black sheep effect': Social categorization, rejection of ingroup deviates, and perception of group variability. *European Review of Social Psychology, 5*(1), 37–68. https://doi.org/10.1080/14792779543000011

McIntosh, P. (1988). *White privilege and male privilege: A personal account of coming to see correspondence through work in women's studies* (working paper no. 189). Wellesley Centers for Women.

Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology, 95*(1), 76–93. https://doi.org/10.1037/0022-3514.95.1.76

Nett, T., Dorrough, A., Jekel, M., & Glöckner, A. (2020). Perceived biological and social characteristics of a representative set of German first names. *Social Psychology, 51*(1), 17–34. https://doi.org/10.1027/1864-9335/a000383

Nicholas, L., & Agius, C. (2018). *The persistence of global masculinism: Discourse, gender and neo-colonial re-articulations of violence*. Springer eBook Collection Social Sciences. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-68360-7.

Nixon, S. A. (2019). The coin model of privilege and critical allyship: Implications for health. *BMC Public Health, 19*(1), 1637. https://doi.org/10.1186/s12889-019-7884-9

Norton, M. I., & Sommers, S. R. (2011). Whites see racism as a zero-sum game that they are now losing. *Perspectives on Psychological Science, 6*(3), 215–218. https://doi.org/10.1177/1745691611406922

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods, 19*, 1–13. https://doi.org/10.1177/1609406919899220

OECD. (2017). *The pursuit of gender equality*. OECD Publishing. https://doi.org/10.1787/9789264281318-en

Oliffe, J. L., Broom, A., Rossnagel, E., Kelly, M. T., Affleck, W., & Rice, S. M. (2020). Help-seeking prior to male suicide: Bereaved men perspectives. *Social Science & Medicine, 261*, 113–173. https://doi.org/10.1016/j.socscimed.2020.113173

Pease, B. (2010). *Undoing privilege: Unearned advantage in a divided world* (1st ed.). Zed Books.

Pease, B. (2021). The rise of angry white men: Resisting populist masculinity and the backlash against gender equality. In C. Noble, & G. F. Ottmann (Eds.), *The challenge of right-wing nationalist populism for social work: A human rights approach* (pp. 55–68). Routledge.

Pinterits, E. J., Poteat, V. P., & Spanierman, L. B. (2009). The white privilege attitudes scale: Development and initial validation. *Journal of Counseling Psychology, 56*(3), 417–429. https://doi.org/10.1037/a0016274

Pinto, I. R., Marques, J. M., Levine, J. M., & Abrams, D. (2010). Membership status and subjective group dynamics: Who triggers the black sheep effect? *Journal of Personality and Social Psychology, 99*(1), 107–119. https://doi.org/10.1037/a0018187

Puryear, C., Steele, L. M., Lawler, J., & Vandello, J. (2019). Reminders of ingroup privilege often threaten identity but rarely change beliefs. *PsyArXiv*. https://psyarxiv.com/gvjkc/download?format=pdf.

Radke, H. R. M., Kutlaca, M., Siem, B., Wright, S. C., & Becker, J. C. (2020). Beyond allyship: Motivations for advantaged group members to engage in action for disadvantaged groups. *Personality and Social Psychology Review, 24*(4), 291–315. https://doi.org/10.1177/1088868320918698

Rasinski, H. M., & Czopp, A. M. (2010). The effect of target status on witnesses' reactions to confrontations of bias. *Basic and Applied Social Psychology, 32*(1), 8–16. https://doi.org/10.1080/01973530903539754

Schmitt, M., Behner, R., Montada, L., Müller, L., & Müller-Fohrbrodt, G. (2000). Gender, ethnicity, and education as privileges: Exploring the generalizability of the existential guilt reaction. *Social Justice Research, 13*, 313–337.

Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science, 8*(4), 379–386. https://doi.org/10.1177/1948550617715068

Shuman, E., Johnson, D., Saguy, T., & Halperin, E. (2018). Threat to the group's image can motivate high identifiers to take action against in-group transgressions. *Personality and Social Psychology Bulletin, 44*(11), 1523–1544. https://doi.org/10.1177/0146167218768800

Shuman, E., van Zomeren, M., Saguy, T., & Knowles, E. (2022). Defend, deny, distance, and dismantle: A measure of how advantaged group members manage their identity. *PsyArXiv*. https://psyarxiv.com/6d4qc/.

Siem, B., Oettingen, M., Mummendey, A., & Nadler, A. (2013). When status differences are illegitimate, groups' needs diverge: Testing the needs-based model of reconciliation in contexts of status inequality. *European Journal of Social Psychology, 43*(2), 137–148. https://doi.org/10.1002/ejsp.1929

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89*(6), 845–851. https://doi.org/10.1037/0022-3514.89.6.845

Steele, C. M. (1988). In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology, (Vol. 21,* pp. 261–302). Academic Press. https://doi.org/10.1016/S0065-2601(08)60229-4.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *34. Advances in Experimental Psychology* (pp. 379–440). Academic Press. https://doi.org/10.1016/S0065-2601(02)80009-0.

The jamovi project. (2021). jamovi (Version 2.2) [Computer software]. https://www.jamovi.org.

Wenzel, M., Woodyatt, L., & McLean, B. (2020). The effects of moral/social identity threats and affirmations on psychological defensiveness following wrongdoing. *British Journal of Social Psychology, 59*(4), 1062–1081. https://doi.org/10.1111/bjso.12378

Supplementary Materials for

Appraisal of male privilege: On the dual role of identity threat and shame in response to confrontations with male privilege

Section A: Study 1

Additional variables not reported in the main manuscript

Male in-group identification. Participants filled out a five-item measure of male in-group identification (adapted from Doosje et al., 1995) on a 7-point scale ranging from 1 = strongly disagree to 7 = strongly agree (e.g., "I identify with other men.", "I am glad to be a man."). The scale had acceptable internal consistency, $\alpha = .77$.

Social Judgment Dimensions. After reading the text, participants completed a battery of scales tapping into social judgment dimensions, measuring competence (competent, intelligent, skilled), sociability (sympathetic, warm, friendly), and morality (honest, trustworthy, sincere) on a 7-point scale with 1 being the negative (e.g., dishonest) and 7 being the positive (e.g., honest) trait (adapted from Leach et al., 2007). Due to recent concerns about the validity of comparing stereotype content means across different target groups, we used a multiple group confirmatory factory analysis to ensure scalar measurement invariance between conditions (Friehs et al., 2022). Scalar measurement equivalence indicates the interpretability of scale means between conditions. In this procedure, two items were deleted (sympathetic, and trustworthy) to reach scalar invariance. The subscales had satisfactory internal consistency ($\alpha_{competence} = .81$, $\rho_{morality} = .92$, $\rho_{sociability} = .73$). The deletion of items did not substantially change results sociability (see the robustness test in the following section). Results are reported below.

Political orientation. Political orientation was measured with a single item raging from 1 (left) to 10 (right).

Social dominance orientation. Social dominance orientation was measured with four items (e.g., "If some groups in the population are superior, this is useful for society"). The scale was adapted from Cohrs and Asbrock (2009). We scattered the scales with additional distractor items to reduce the demand characteristics of the study. The scale was used to estimate the confirmatory validity of meritocratic threat (see below).

Differences in social judgement dimensions between conditions

Author evaluation differed on all three dimensions of social judgements: competence, $F(2,649) = 8.69$, $p < .001$, $\eta^2p = .03$, sociability, $F(2,650) = 33.6$, $p < .001$, $\eta^2p = .09$, and morality,

$F(2,651) = 10.4$, $p < .001$, $\eta^2p = .03$. Tukey-corrected post-hoc tests revealed that the female confronter was evaluated significantly less favorable than the male author on competence, $t(649) = -2.65$, $p = .02$, $d = -.26$, 95% CI [-.45 -.07], sociability, $t(650) = -7.91$, $p < .001$, $d = -.76$, 95% CI [-.96 -.57], and morality, $t(651) = -3.92$, $p < .001$, $d = -.38$, 95% CI [-.57 -.19].

Measurement invariance procedure. We computed confirmatory factor analysis with robust maximum likelihood estimation to establish the baseline model with three factors (competence, sociability, morality). We ran this analysis multiple times, each after deletion of the item with the highest modification index (> 4) regarding factor loadings. In this way, the items "trustworthy" and "likebale" were deleted. The baseline model reached good model fit, $\chi^2(11) = 22.50$, $p = .02$, CFI = 1.00, SRMR = .01, RMSEA = .04, 95% CI [.02 .06]. Subsequently we tested measurement invariance via multi-group confirmatory factor analysis with stepwise introduction of equality constraints across conditions. Constraints score tests between each iteration indicated whether the additional constraint significantly decreased model fit. The constraint score test was non-significant between the configural (no constraints) and metric invariance model (factor loadings equal), $\chi^2(8) = 6.40$, $p = .60$, and between the metric and scalar invariance (intercepts equal) step, $\chi^2(22) = 25.14$, $p = .29$. As such, the model reached scalar invariance. Scalar measurement invariance assures that comparing means between different groups is meaningful (van de Schoot et al., 2012).

Robustness test. The pattern of results did not depend on the use of the original three-item morality and sociability scales or their measurement invariant two-item versions. An ANOVA with the original morality scale as DV was significant, $F(2,651) = 16.91$, $p < .001$, $\eta^2_p = .05$, and Tukey-corrected post-hoc tests revealed that the female confronter received lower morality ratings compared to the male confronter, $t(651) = -5.45$, $p < .001$, $d = -.53$, 95% CI [-.72 -.33], and compared to the work group behind the information, $t(651) = -4.50$, $p < .001$, $d = -.43$, 95% CI [-.24 -.62]. There was no difference between the male confronter and the control condition, $t(651) = 1.01$, $p = .58$, $d = .10$, 95% CI [-.09 .28].

An ANOVA with the original sociability scale as DV was significant, $F(2,650) = 31.81$, $p < .001$, $\eta^2_p = .09$, and Tukey-corrected post-hoc tests revealed that the female confronter received lower sociability ratings compared to the male confronter, $t(650) = -7.40$, $p < .001$, $d = -.71$, 95% CI [-.91 -.52], and compared to the work group behind the information, $t(650) = -6.29$, $p < .001$, $d = -.60$, 95% CI [-.79 -.41]. There was no difference between the male confronter and the control condition, $t(650) = 1.18$, $p = .46$, $d = .11$, 95% CI [-.08 .30].

Meritocratic threat scale validation with SDO

We validated the new meritocratic threat scale in two steps. First, an exploratory factor analysis with meritocratic and image threat using maximum likelihood extraction, promax rotation, and parallel analysis extracted two factors, which were moderately correlated (r = .45). Second, we did a simple test of construct validity by correlating meritocratic threat and image threat with social dominance orientation (SDO). SDO has been found in previous research to correlate highly with support for meritocratic policies (Kim & Berry, 2015). We expected and found that SDO more strongly correlates with meritocratic threat (r = .23, $p < .001$) than with group-image threat (r = .10, $p = .01$).

Correlation matrix

Table 1. Correlation matrix with relevant dependent variables

		1	2	3	4	5	6	7	8
1 Group-Image Threat		—							
2 Meritocratic Threat		.377 ***	—						
	df	652							
3 Privilege Awareness		-.195 ***	-.542 ***	—					
	df	652	652						
4 Privilege Remorse		-.052	-.329 ***	.455 ***	—				
	df	652	652	652					
5 Auth. Evaluation		-.298 ***	-.393 ***	.470 ***	.367 ***	—			
	df	636	636	636	636				
6 Competence		-.161 ***	-.338 ***	.491 ***	.353 ***	.548 ***	—		
	df	650	650	650	650	634			
7 Sociability		-.198 ***	-.324 ***	.333 ***	.224 ***	.553 ***	.552 ***	—	
	df	651	651	651	651	635	650		
8 Morality		-.160 ***	-.349 ***	.482 ***	.263 ***	.524 ***	.636 ***	.759 ***	—
	df	652	652	652	652	636	650	651	

Note. * p < .05, ** p < .01, *** p < .001.

Section B: Study 2

Differences in social judgement between conditions

While the female author did receive a lower evaluation on the sociability dimension, $t(183.87) = -4.21$, $p < .001$, $d = -0.62$, 95% CI [-0.92 -0.32], there was on difference on the morality, $t(183.91) = 0.13$, $p = .55$, $d = 0.02$, 95% CI [-.27 .31], and competence, $t(180.92) = 0.83$, $p = .80$, $d = 0.12$, 95% CI [-0.17 0.41] dimensions.

Analysis of measurement invariance was conducted similarly to Study 1. For the baseline model the items "trustworthy", "likeable", and "intelligent" were deleted. The baseline model reached good model fit, $\chi^2(6) = 12.33$, $p = .06$, CFI = .99, SRMR = .02, RMSEA = .08, 95% CI [.00 .14]. The constraint score test was non-significant between the configural (no constraints) and metric invariance model (factor loadings equal), $\chi^2(3) = 3.57$, $p = .31$, and between the metric and scalar invariance (intercepts equal) step, $\chi^2(9) = 11.35$, $p = .25$. As such, the model reached scalar invariance. The pattern of results did not depend on the use of the original three-item morality and sociability scales or their measurement invariant two-item versions. A one-sided Welch's *t*-Test with the original morality scale found no difference between both conditions, $t(183.87) = -0.01$, $p = .50$, $d = -.002$, 95% CI [-.29 .29]. The female (versus male) author received lower ratings on sociability, $t(183.96) = -3.65$, $p < .001$, $d = -.54$, 95% CI [-.83 - .24].

Outlier handling

As pre-registered, we ran an analysis of univariate outliers based on box plots with all dependent variables. There were $n = 29$ outliers on moral threat (> 1.5 IQD). To check for the robustness of our results, we tested whether moral threat was still affected by condition after the deletion of outliers. After exclusion of univariate outliers, shame did not differ significantly between the female confronter ($M = 1.32$, $SD = 0.60$) and the male confronter condition ($M = 1.44$, $SD = 0.75$), $t(138.58) = -1.07$, $p = .14$, $d = -.17$, 95% CI [-.49 .14].

Detailed results of codes between conditions

Table 3. Frequency of reactions to the confrontation based on confronter gender.

Reaction	Code frequency	Female Confronter Sum (% of n = 95)	Male Confronter Sum (% of n = 91)
Acceptance	331	167 (30.6)	164 (32.8)
General acceptance	105		
Change talk	70		
Limited acceptance	81		
Supplemented acceptance	75		
Justification*	260	154 (34.9)	106 (26.3)
Rejection of responsibility	165		
Powerless individual	16		
Legitimization	47		
Overgeneralization	32		
Rejection	230	112 (29.0)	118 (26.9)
Single cases	58		
Mixed evidence	92		
Reverse discrimination	40		
Whataboutism	31		
Unspecific rejection	9		

Backlash*	112	81 (22.6)	50 (16.1)
Victim blaming	13		
Devaluation	99		

Note. * one-sided t-test $p < .05$. Significance tests were only performed on the category level to avoid Type I error inflation.

Emotional reactions to the two confronters

Two-sided Welch's T-Tests were computed across the two conditions with emotional reactions as dependent variables. Participants reported more shame in the male confronter (M = 2.01, SD = 1.35) than in the female confronter condition (M = 1.64, SD = 1.06), $t(170.53) = -2.06$, p = .02, d = -.30, and shame was positively related to text, b = 6.11, p = .004, β = .21, 95% CI [.07 .36], and author evaluation, b = 5.11, $p < .001$, β = .24, 95% CI [.10, .38]. Shame was also associated with less frequency of rejecting male privilege, b = -0.37, p = .003, β = -.22, 95% CI [-.36 -.08], but was not significantly related to acceptance, b = 0.21, p = .07, β = .14, 95% CI [-.009 .28] and justification, b = -0.15, p = .13, β = -.11, 95% CI [-.26 .03].

Regarding the additionally measured emotional responses, gender of the confronter only significantly affected the level of gratitude, with the male confronter triggering higher levels of gratitude (M = 4.50, SD = 0.91) than the female confronter (M = 4.19, SD = 1.16), $t(170.32) = 2.01$, p = .047, d = .30. All means and standard deviations are presented in Table 4.

Table 4. Descriptive information regarding the emotions to the confrontations

	Condition	Annoyance	Anger	Affliction	Sadness	Pride (reversed)	Gratitude (reversed)	Confusion
Mean	female	3.10	2.24	2.67	2.39	4.60	4.50	2.20
	male	2.96	2.30	2.31	2.30	4.56	4.19	2.12
SD	female	1.30	1.27	1.40	1.39	0.83	0.91	1.24
	male	1.35	1.30	1.24	1.28	0.79	1.16	1.27

Test of shame versus guilt as predictors of privilege attitude and confronter evaluation

Some works have identified guilt as a specifically "moral emotion" (Fitouchi et al., 2021), with some potential to motivate restitution (Leach et al., 2002). We therefore also investigated whether shame and guilt have unique effects on confronter evaluation and privilege attitude based on multiple linear regressions with both emotions as covariates. In this analysis we found that the effect of shame on both outcomes disappeared when controlling for guilt. Specifically, we found that guilt had a unique effect on privilege attitude, b = 12.40, 95% CI [0.20 0.52], β = .36, $p < .001$, but shame

did not, b = 0.64, 95% CI [-0.14 0.18], β = .02, p = .78, and that guilt had a unique effect on author evaluation, b = 7.27, 95% CI [0.12 0.45], β = .29, p < .001, but shame did not, b = 1.86, 95% CI [-0.08 0.25], β = .09, p = .29.

Correlation matrices of all dependent variables

Table 5 presents correlations between all variables reported in the manuscript. Table 6 presents additional correlations between shame and code frequencies.

Table 5. Correlation matrix with relevant dependent variables

		1	2	3	4	5	6
1	Shame	-					
2	Privilege Attitude	.21**	-				
3	Confronter Evaluation	.24***	.76***	-			
4	Competence	.26***	.69***	.69***	-		
5	Morality	.06	.56***	.53***	.61***	-	
6	Sociability	.14	.58***	.75***	.50***	.59***	-

Note. ** p < .01, *** p < .001

Section C: Study 3

Detailed report on the pretest including scale construction

N = 49 adult male participants completed the online pretest via prolific academic. Mean age was *M* = 34.78 (*SD* = 11.41). We measured group-image and moral threat with three items, image and moral shame with two items and privilege attitude polarities across eight items. The response scales were unchanged in the final study.

To test the validity of the adapted/new scales for moral threat, image shame, moral shame, and privilege attitude polarities, we computed correlations and exploratory factor analyses. The original scales together with changes and decision criteria are summarized in Table 7.

Table 7. Original scale with changes and decision criteria

Original Item	Changes	Decision criterion
Moral Threat		
... that men have a moral deficit. ... that men are morally failing.		
... that men are flawed in some way.	Deleted	Low correlation with rest of scale (< .50)
Image Shame		
I am ashamed because men are portrayed negatively I am ashamed of the negative impression men give		
Moral Shame		
I feel ashamed because men are not the "good guys".	I am ashamed of the undeserved advantages men have	Similar correlation between this and other moral and image shame items (all r .40-.50)
I feel ashamed because of the unjustified preferential treatment of men		
Privilege attitude polarities		
Men generally have it harder than women/Men generally have it easier than women	Deleted and used similar item for privilege awareness scale	To retain comparability with Study 1
Feminism goes too far/Feminism does not go far enough		

Men are wrongly/rightly blamed on the issue of male privilege		
The issue of male privilege is only for male bashing./The issue of male privilege is every man's business.	Deleted and used similar item for privilege awareness scale	To retain comparability with Study 1
I have mainly disadvantages/advantages in my life because I am a man		
Personally, I have not benefited at all/a lot from being a man		
I owe my personal success mainly to my own efforts./I owe my personal success mainly to my male privilege.	Deleted	Low mean (*M* = 2.71, *SD* = 1.67)
I don't want/want to take responsibility for dismantling male privilege. I have no/great interest in learning more about the topic of male privilege		

Factor analyses for threat, shame, and privilege attitudes

Threat

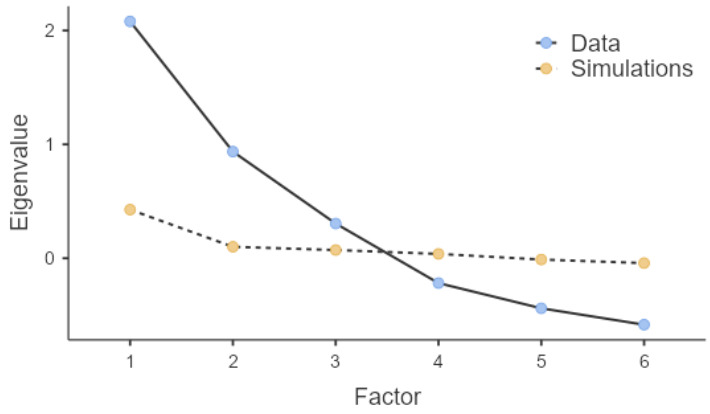
Results of the exploratory factor analysis across the six threat items are presented in Table 8. Parallel analysis was used for factor extraction (see Figure 1).

Table 8. Exploratory factor analysis across threat items

	Factor		
	Merit	Moral	Image
… I had the impression that men look bad because of their privilege			0.75
… I had the impression that advantaging men makes other people have a negative image of my gender			0.77
… I had the impression that men have a moral deficit		0.88	
… I had the impression that men are morally failing.		0.90	
… that my successes are due to my gender	0.95		
… that my social status is based on my gender	0.95		

Note. 'Principal axis factoring' extraction method was used in combination with 'oblimin' rotation

Figure 1. Parallel analysis plot for EFA across threat items



Shame

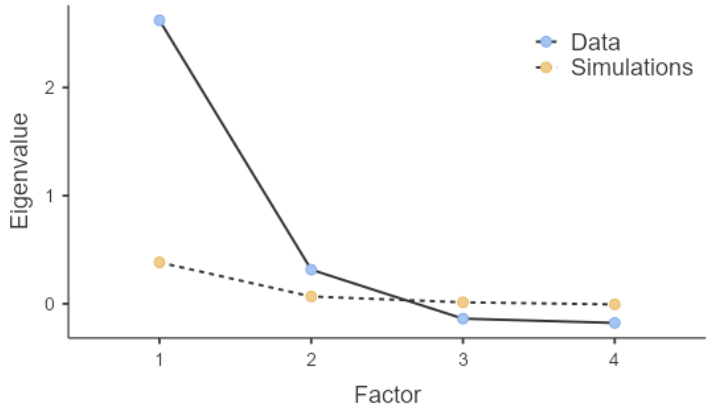
Results of the exploratory factor analysis across the four shame items are presented in Table 9. Parallel analysis was used for factor extraction (see Figure 2).

Table 9. Exploratory factor analysis across shame items

	Factor	
	Moral	Image
I am ashamed because men are portrayed negatively		0.89
I am ashamed of the negative impression men give		0.83
I am ashamed of the undeserved advantages men have	0.92	
I feel ashamed because of the unjustified preferential treatment of men	0.92	

Note. 'Principal axis factoring' extraction method was used in combination with 'oblimin' rotation

Figure 2. Parallel analysis plot for EFA across shame items



Awareness

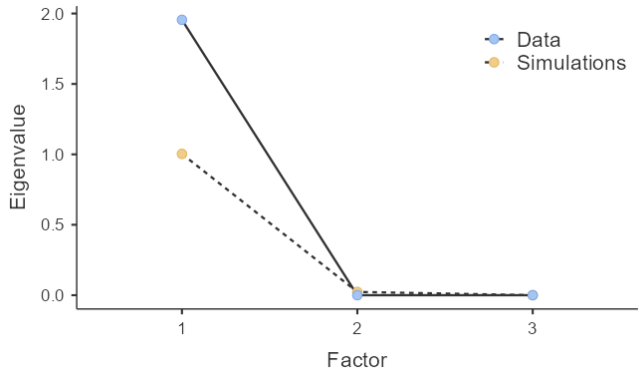
Results of the exploratory factor analysis across the three privilege awareness items are presented in Table 10. Parallel analysis was used for factor extraction (see Figure 3).

Table 10. Exploratory factor analysis across privilege awareness items

	Factor
	Awareness
In general, men have it easier than women	0.82
The structure of our social system promotes male privilege	0.92
Everyone has the same opportunities, so these so-called male privileges are just male bashing (reversed)	0.67

Note. 'Principal axis factoring' extraction method was used in combination with 'oblimin' rotation

Figure 3. Parallel analysis plot for EFA across privilege awareness items



Privilege attitudes polarities

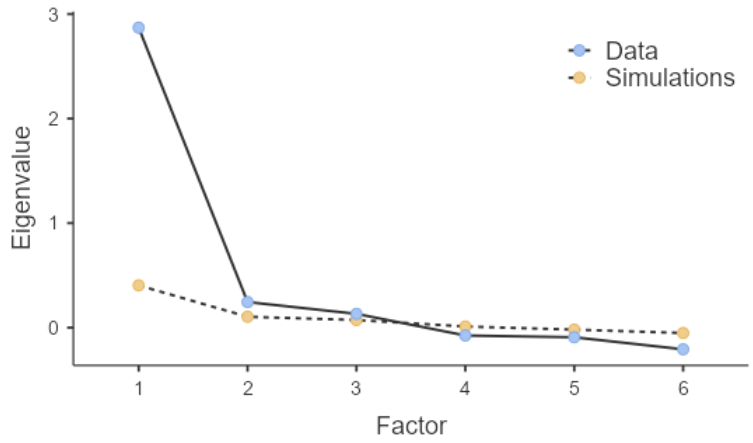
Results of the exploratory factor analysis across the six privilege attitude polarities items are presented in Table 11. Parallel analysis was used for factor extraction (see Figure 4).

Table 11. Exploratory factor analysis across privilege attitudes polarities

	Factor		
	Responsibility	Feminism	Status
Feminism goes too far/Feminism does not go far enough		0.64	
Men are wrongly/rightly blamed on the issue of male privilege		0.83	
I have mainly disadvantages/advantages in my life because I am a man			0.64
Personally, I have not benefited at all/a lot from being a man			0.67
I don't want/want to take responsibility for dismantling male privilege	0.81		
I have no/great interest in learning more about the topic of male privilege	0.67		

Note. 'Principal axis factoring' extraction method was used in combination with 'oblimin' rotation

Figure 4. Parallel analysis plot for EFA across privilege attitudes polarities



Descriptive comparison of demographic data between conditions

A descriptive comparison along age, political orientation, subjective social status, migration background, and nationality and is presented in Table 12.

Table 12. Descriptive comparison of demographic data between conditions

	Condition	N	Mean	SD
Age	Female	211	38.10	15.75
	Male	190	36.90	14.96
	Information	222	37.26	14.60
Political Orientation	Female	211	6.20	1.62
	Male	190	5.98	1.59
	Information	222	5.81	1.65
Subjective Social Status	Female	211	44.08	18.70
	Male	190	43.03	20.20
	Information	222	43.89	20.19
		Count	% of condition	
Migration Background	Female	62	30.38%	
	Male	73	38.42%	
	Information	91	40.99%	
German Nationality	Female	191	90.52%	
	Male	173	91.05%	
	Information	201	90.54%	

Effects on anger

There was a surprisingly strong effect between conditions on anger, $F(2,620) = 27.37$, $p < .001$, $\eta_p^2 = .08$. The female confronter elicited more anger than the male confronter, $t(620) = 3.95$, $p_{tukey} < .001$, $d = .40$ (95% CI [0.20 0.59]), and the information, $t(620) = 7.39$, $p_{tukey} < .001$, $d = .71$ (95% CI [0.52 0.90]). Also, the male confronter triggered more anger than the information, $t(620) = 3.20$, $p_{tukey} = .004$, $d = .32$ (95% CI [0.12 0.51]).

Detailed results of the paths in the appraisal-based model

Detailed results for each individual path in the appraisal-based model are presented in Table 13.

Table 13. Information on all paths in the appraisal-based model

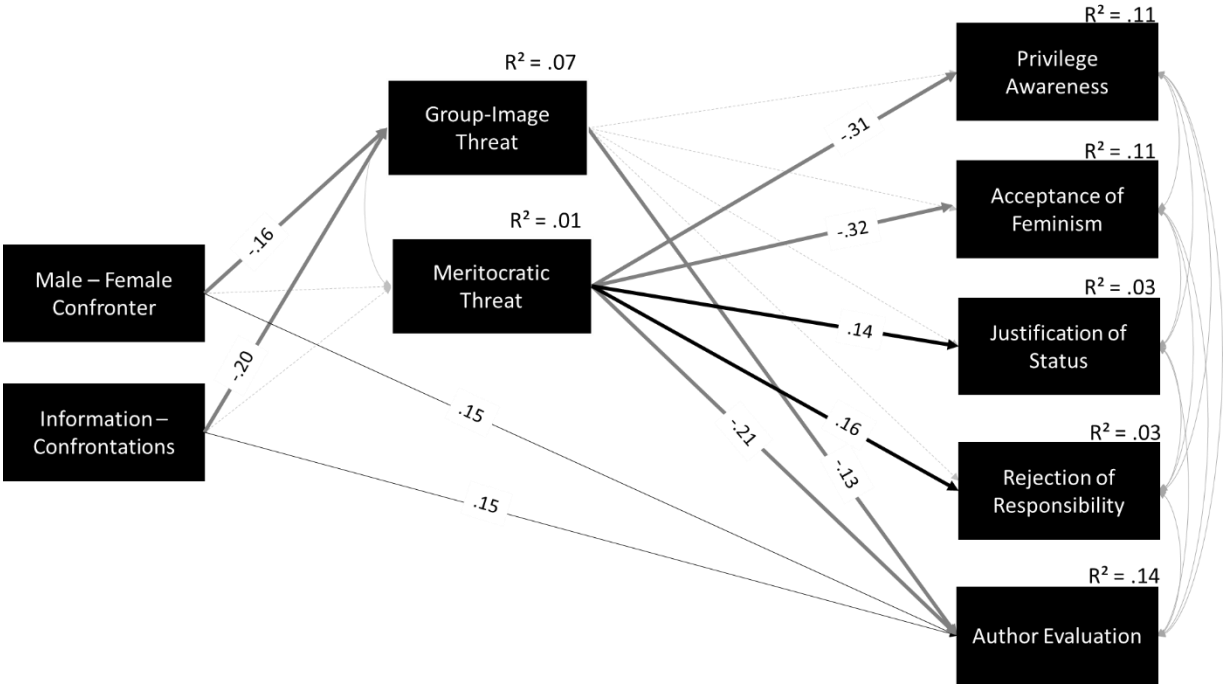
			95% CI			
Dependent	Predictor	b	LCI	UCI	β	p
Image Threat	Male-Female	-0.64	-0.92	-0.35	-.16	< .001
Image Threat	Information-Confrontations	-0.65	-0.91	-0.40	-.20	< .001

Moral Threat	Male-Female	-0.39	-0.72	-0.06	-.09	.02
Moral Threat	Information-Confrontations	-0.38	-0.65	-0.11	-.11	.01
Image Shame	Image Threat	0.30	0.20	0.40	.21	< .001
Moral Shame	Moral Threat	0.27	0.18	0.36	.20	< .001
Awareness	Image Shame	-0.21	-0.26	-0.15	-.30	< .001
Awareness	Moral Shame	0.40	0.34	0.46	.58	< .001
Acceptance	Image Shame	-0.17	-0.22	-0.11	-.24	< .001
Acceptance	Moral Shame	0.42	0.37	0.48	.63	< .001
Justification	Image Shame	0.08	0.03	0.14	.15	.001
Justification	Moral Shame	-0.21	-0.26	-0.15	-.36	< .001
Rejection	Image Shame	0.10	0.04	0.16	.15	< .001
Rejection	Moral Shame	-0.42	-0.47	-0.36	-.61	< .001

3D Model configuration

The model with meritocratic threat and group-image threat as parallel mediators had good model fit, $\chi^2(8) = 8.16$, $p = .42$, CFI = 1.00, RMSEA = 0.006 ($p = .96$, 95% CI [.00 .05]), SRMR = .02. It is summarized in Figure 5. Indirect effects are presented in Table 14.

Figure 5. Path model based on the 3D model configuration of mediators



Note. Estimates are standardized. All paths with estimates are significant $p < .05$. Dashed arrows represent negative effects. Narrow arrows represent the estimated direct effects between condition

and author evaluation. Curved arrows represent estimated residual covariance between the mediators.

Table 14. Indirect effects

	Path	Outcome	*b*	LCI	UCI	β	*p*
Male-Female confronter contrast							
IE1	Merit	Awareness	0.04	-0.06	0.13	.01	.45
IE2	Merit	Acceptance	0.04	-0.06	0.14	.01	.45
IE3	Merit	Justification	-0.01	-0.05	0.02	.00	.46
IE4	Merit	Rejection	-0.02	-0.07	0.03	-.01	.46
IE5	Merit	Evaluation	0.43	-0.68	1.53	.01	.45
IE6	Image	Awareness	0.03	-0.03	0.09	.01	.34
IE7	Image	Acceptance	0.00	-0.05	0.06	.00	.90
IE8	Image	Justification	-0.04	-0.09	0.01	-.01	.14
IE9	Image	Rejection	0.00	-0.06	0.06	.00	.97
IE10	Image	Evaluation	1.38	0.21	2.55	.02	.02
Information-Confrontations contrast							
IE11	Merit	Awareness	0.07	0.02	0.15	.02	.09
IE12	Merit	Acceptance	0.07	0.01	0.15	.02	.09
IE13	Merit	Justification	-0.03	-0.03	0.01	-.01	.14
IE14	Merit	Rejection	-0.04	-0.04	0.01	-.01	.14
IE15	Merit	Evaluation	0.79	0.18	1.76	.01	.11
IE16	Image	Awareness	0.03	-0.07	0.09	.01	.35
IE17	Image	Acceptance	0.00	-0.08	0.06	.00	.90
IE18	Image	Justification	-0.04	0.01	0.01	-.02	.15
IE19	Image	Rejection	0.00	0.01	0.06	.00	.97
IE20	Image	Evaluation	1.41	-0.95	2.59	.03	.02

Test of the robustness of shame dimensions for the inclusion of guilt

In Study 2, the effects of shame on our outcomes depended on the correlation between shame and guilt. Allpress et al. (2014) also observed the role of guilt in conjunction with the image and moral dimensions of shame and found less consistent effects for guilt than for these shame dimensions. As such, we assumed that our simple measure of shame masked the positive and negative effects of image shame and moral shame, respectively, and that therefore investigating both dimensions of shame separately will draw out the unique effect of shame. We therefore tested whether both shame dimensions affect privilege attitudes independently of guilt. A hierarchical regression with guilt and both dimensions of shame as predictors of privilege awareness was conducted. Results indicated that moral shame and image shame explained variance in privilege awareness over and above the effect of guilt and condition. Hence, moral and image shame seem to be more proximal predictors of privilege awareness than guilt, which is in line with previous research

(Allpress et al., 2014). The results of the hierarchical regression on privilege awareness including models for condition, guilt, and both dimensions of shame are summarized in Table 15.

Table 15. Hierarchical linear regression on privilege awareness

	Model 1		Model 2		Model 3	
	b	*p*	b	*p*	b	*p*
Intercept	4.42	< .001	4.42	< .001	4.42	< .001
Male – Female	0.27	.08	0.20	.19	0.23	.10
Information – Confrontations	0.24	.07	0.18	.15	0.14	.21
Guilt	-	-	0.21	< .001	0.06	.08
Image Shame	-	-	-	-	-0.21	< .001
Moral Shame	-	-	-	-	0.36	< .001

Correlation matrix

Table 16. Correlation matrix across all dependent variables in Study 3

	1	2	3	4	5	6	7	8	9	10	11	12	
1 Image Threat	—												
2 Moral Threat	.40 ***	—											
3 Merit Threat	.29 ***	.15 ***	—										
4 Image Shame	.21 ***	.32 ***	.33 ***	—									
5 Moral Shame	.08 *	.33 ***	.08	.63 ***	—								
6 Guilt	.07	.27 ***	.13 **	.52 ***	.66 ***	—							
7 Anger	.32 ***	.03	.53 ***	.25 ***	-.08	-.05	—						
8 Awareness	-.14 ***	.10 *	-.32 ***	.08 *	.42 ***	.29 ***	-.54 ***	—					
9 Acceptance	-.10 *	.18 ***	-.32 ***	.17 ***	.49 ***	.42 ***	-.52 ***	.69 ***	—				
10 Justification	.12 **	-.06	.16 ***	-.09 *	-.28 ***	-.27 ***	.32 ***	-.51 ***	-.49 ***	—			
11 Rejection	.04	-.19 ***	.16 ***	-.25 ***	-.53 ***	-.44 ***	.36 ***	-.54 ***	-.66 ***	.48 ***	—		
12 Author Evaluation	-.24 ***	.06	-.26 ***	.06	.32 ***	.32 ***	-.56 ***	.47 ***	.55 ***	-.35 ***	-.49 ***	—	
13 Text Evaluation	-.22 ***	.06	-.27 ***	.08 *	.33 ***	.32 ***	-.57 ***	.51 ***	.57 ***	-.37 ***	-.51 ***	.86 ***	—

Note. * p < .05, ** p < .01, *** p < .001, *N* = 623 in all cells.

References

Allpress, J. A., Brown, R., Giner-Sorolla, R., Deonna, J. A., & Teroni, F. (2014). Two faces of group-based shame: Moral shame and image shame differentially predict positive and negative orientations to ingroup wrongdoing. *Personality & Social Psychology Bulletin*, *40*(10), 1270–1284. https://doi.org/10.1177/0146167214540724

Cohrs, J. C., & Asbrock, F. (2009). Right-wing authoritarianism, social dominance orientation and prejudice against threatening and competitive ethnic groups. *European Journal of Social Psychology*, *39*(2), 270–289. https://doi.org/10.1002/ejsp.545

Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology*, *31*(5), 410–436. https://doi.org/10.1006/jesp.1995.1018

Fitouchi, L., André, J.-B., & Baumard, N. (2021). *Are there really so many moral emotions? Carving morality at its functional joints*. https://psyarxiv.com/2phzt/

Friehs, M.-T., Kotzur, P. F., Böttcher, J., Zöller, A.-K. C., Lüttmer, T., Wagner, U., Asbrock, F., & van Zalk, M. H. W. (2022). Examining the structural validity of stereotype content scales – A preregistered re-analysis of published data and discussion of possible future directions. *International Review of Social Psychology*, *35*(1), 1. https://doi.org/10.5334/irsp.613

Kim, A., & Berry, C. M. (2015). Individual differences in social dominance orientation predict support for the use of cognitive ability tests. *Journal of Personality*, *83*(1), 14–25. https://doi.org/10.1111/jopy.12078

Leach, C. W., Snider, N., & Iyer, A. (2002). "Poisoning the conscious of the fortunate": The experience of relative advantage and support for social equality. In I. Walker & H. J. Smith (Eds.), *Relative Deprivation* (pp. 136–163). Cambridge University Press.

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

Manuscript 2

Addressing the “Why” in White privilege: The limits of mere evidence confrontations and the potential of contextualizing White privilege in national colonial history

Frank Eckerle, Carmen S. Lienen, Maja Kutlaca, Patrick F. Kotzur, & J. Christopher Cohrs

Studies found that when confronted with evidence of White privilege, White individuals claim life hardships to distance themselves from it. However, these effects were only barely significant, and samples were predominantly from the US, which is a unique context for White privilege. Merely presenting evidence of White privilege also overlooks its socio-historical embedding in colonialism. Drawing from collective memory research, we expected contextualizing White privilege in national colonial history to trigger moral concerns. None of three experiments with White Germans (N = 646) replicated the effect on life hardships. Instead, the contextualized confrontation activated moral motives for social justice among White German students and White Germans with a migration background. We highlight the need for theoretical models that explicitly consider how cultural contexts shape reactions to White privilege even among "Western" nations. We discuss the social change potential of framing White privilege as a systemic rather than individualized issue.

Keywords: White Privilege, Life Hardships, Social Justice Morality, Colonialism, Collective Memory

Addressing the "Why" in White Privilege: The limits of mere evidence confrontations and the potential of contextualizing White privilege in national colonial history

Frank Eckerle[1], Carmen S. Lienen[1], Maja Kutlaca[2], Patrick F. Kotzur[2], & J. Christopher Cohrs[1]

[1] Philipps-University Marburg, Department of Psychology, Gutenbergstr. 18, 35032 Marburg, Germany

[2] Durham University, Department of Psychology, Upper Mountjoy, South Rd, Durham DH1 3LE, United Kingdom

Correspondence concerning this article should be addressed to Frank Eckerle, Gutenbergstr. 18, FB04 Psychologie, 35032 Marburg, Germany. Email: frank.eckerle@uni-marburg.de

Addressing the "Why" in White privilege: The limits of mere evidence confrontations and the potential of contextualizing White privilege in national colonial history

Studies found that when confronted with evidence of White privilege, White individuals claim life hardships to distance themselves from it. However, these effects were only barely significant, and samples were predominantly from the US, which is a unique context for White privilege. Merely presenting evidence of White privilege also overlooks its socio-historical embedding in colonialism. Drawing from collective memory research, we expected contextualizing White privilege in national colonial history to trigger moral concerns. None of three experiments with White Germans (N = 646) replicated the effect on life hardships. Instead, the contextualized confrontation activated moral motives for social justice among White German students and White Germans with a migration background. We highlight the need for theoretical models that explicitly consider how cultural contexts shape reactions to White privilege even among "Western" nations. We discuss the social change potential of framing White privilege as a systemic rather than individualized issue.

Keywords: White Privilege, Life Hardships, Social Justice Morality, Colonialism, Collective Memory

Social privileges are unearned advantages conferred via certain identifiers, such as being White, male, heterosexual, middle-class, or able-bodied. People with many such privileges are often respected more, given more opportunities, and less likely to become targets of prejudice and discrimination (see for example McIntosh, 1988). Because these advantages often go unchecked, people are not aware of their social privileges and, when confronted, tend to justify their status by highlighting individual effort (Phillips & Jun, 2021). Previous work argues that social privilege should be understood as the flipside of discrimination, as both privilege and discrimination originate from systems of inequality (e.g., racism, sexism) and connected practices of oppression that have deep historical and cultural roots (Bergkamp, Martin, & Olson, 2022; Nixon, 2019). White privilege, for example, can be traced back to the invention of *race*, which justified exploitation and atrocities in the name of colonialism (Bonds & Inwood, 2016). To dismantle White privilege, it is therefore not enough to make people aware of privileges they enjoy today. What is needed is an understanding of the systemic roots of privilege and a motivation to change the structural forces that perpetuate it (Täuber & Moughalian, 2022). We propose that carefully designed interventions that

highlight the systemic nature of privilege provide one pathway toward that goal (Case et al., 2012).

Past research documents that confrontations with White privilege often trigger defensive responses (Knowles et al., 2014; Lowery et al., 2012; Murdoch & McAloney-Kocaman, 2019; Phillips & Lowery, 2015; Unzueta & Lowery, 2008), such as denying inequality or justifying privilege through merit (Knowles et al., 2014). Whereas some interventions found positive effects, such as increased acceptance that one personally profits from privilege and motivation to address the inequality (Case et al., 2014; Ehrke et al., 2020; Leach et al., 2006), others found that White privilege confrontations rather tend to exacerbate previously held attitudes toward racial inequality instead of changing them (Brad et al., 2019; Puryear et al., 2019; Strupp-Levitsky, 2020). What can explain these different outcomes? We propose that the answer lies in the use of different intervention styles. For instance, defensive effects of privilege interventions have been found after confronting participants with mere evidence of social privilege, such as short paragraphs about how "White Americans enjoy many privileges that Black Americans do not" (Phillips & Lowery, 2015, p. 13). In contrast, the positive effects stem from more vivid and thorough reflections on privilege (Case et al., 2014), interventions embedded in diversity trainings (Ehrke et al., 2020), and through witnessing outgroup discrimination (Uluğ & Tropp, 2021). What these effective interventions have in common is that they aim for a deeper understanding of privilege and of its relation to inequality. We suggest that providing such context to privilege confrontations can activate a deliberation on the immorality of injustice, thereby raising the intervention's effectiveness.

Explaining defensive and constructive responses: The 3D model

Previous research suggests that people's reaction to confrontations with their (White) privilege can be understood as a function of a threatened social identity (Knowles et al., 2014): People identify with and draw self-esteem from their membership in salient social groups, which motivates people to maintain a positive evaluation of their ingroups (Abrams & Hogg, 1990; Tajfel & Turner, 1979). In these terms, White privilege interventions can highlight (i.e., make salient) and threaten White identity, motivating responses to retain a positive self-evaluation such as the three "ds" summarized in the 3D model: *denying* the existence of privilege or injustice, *distancing* the self from Whiteness or from White privilege (e.g., by claiming life hardships), or *dismantling* the system of inequality that perpetuates White privilege (Knowles et al., 2014). The 3D model theorizes that whether responses to privilege interventions are defensive or constructive depends on the nature of identity threat.

Specifically, it considers meritocratic threat (i.e., threat to the justification of status via individual effort) to be a motivator of defensive responses and group-image threat (i.e., threat to the ingroup's standing in the eyes of others) a motivator of constructive responses.

The negative effect of meritocratic threat has been well documented in previous research in the US and the UK. This research demonstrated that confrontations with mere evidence of White privilege trigger claims of life hardships, leading to less acceptance of White privilege (i.e., distancing), less belief in the existence of White privilege (i.e., denial), and less support for pro-egalitarian policies (i.e., dismantle; Murdoch & McAloney-Kocaman, 2019; Phillips & Lowery, 2015). This finding needs further evaluation, however, because the way in which racism and White privilege is discussed differs strongly between European and US-American contexts (Ball et al., 2022). Additionally, the experiments reported in both publications relied on empty control conditions (i.e., participants jumped right to the items instead of completing an unrelated task) and produced relatively small effects. Our goal is therefore to replicate these findings beyond the specific cultural context they originated in and to address the recent criticism regarding the over-estimation of effects sizes and use of less adequate (i.e., empty) control conditions (Lane et al., 2021; Schäfer & Schwarz, 2019). Following previous research, we expect that a mere evidence confrontation (compared to a control condition) will lead to more claims of life hardships (H1a) and that life hardships will then explain the negative effect of the mere evidence confrontation on acceptance of personal privilege and solidarity-based collective action intentions (H1b).

The assumed constructive effect of group-image threat has not received much attention (but see e.g., Shuman et al., 2018) and there are some indications that concerns for the ingroup's image do not lead to more acceptance as predicted by 3D model, but rather trigger defensive reactions. For example, British people who were reminded of atrocities done by their British ingroup during the Iraq war reported two facets of shame: image shame (i.e., threat to social image) and moral shame (i.e., threat to moral essence). Only image shame was positively related to anger and avoidance-tendencies. In contrast, moral shame led to more positive responses, such as support for apologies and compensation (Allpress et al., 2014). The positive impact of moral (versus image) appraisal of ingroup privilege has also recently been documented in the context of gender inequality (Eckerle et al., 2023). As such, we expect moral reflection on inequality to be the key driver of positive responses to White privilege confrontations. We propose that contextualizing privilege might increase the likelihood of moral reflection.

The potential benefits of a contextualized, post-colonial White privilege intervention

In the case of White privilege, contextualizing an intervention means linking the rise of the ideology of White supremacy and the extraction of work power and resources during colonialism to systemic advantages enjoyed by White people today. Popular books about White privilege published in Germany usually position the origin of White privilege in colonialism as well (e.g., Ogette, 2023; Sow, 2009). A contextualized confrontation therefore confronts with (the origin of) racist practices embedded in Western society that led to White people's advantage, and focuses on oppression *and* privilege (Lucal, 1996). Confrontations with White privilege make an individual's identity as White salient (Knowles et al., 2014). However, a contextualized confrontation emphasizes more strongly the role of systemic inequality rooted in racism. As such, a contextualized intervention shifts the focus from individual to collective responsibility. This means that in contrast to a mere evidence confrontation, a contextualized confrontation does not implicate the individual directly as much and may instead activate moral concerns about systemic inequality. Specifically, moral concerns may be activated when White privilege is discussed in the context of historical injustices that were facilitated by an ideology of White supremacy (i.e., colonial exploitation). We argue that by providing this important context to White privilege confrontations, participants may see an opportunity for moral redemption by engaging in constructive responses such as political solidarity. This follows the idea that when the invisible hegemonic forces which perpetuate White privilege are uncovered, privilege is more likely to be accepted (Bergkamp, Olson, & Martin, 2022).

Connecting the past to the present

Even though White Europeans living today have not participated in past colonial practices themselves, they may still identify with and feel responsible for their nation's collective history. Building on the notion that individuals draw part of their self-concept from the groups they belong to (Abrams & Hogg, 1990; Tajfel & Turner, 1979), collective memory research has suggested that people can feel "guilty by association" and be motivated to compensate for past ingroup behavior when the past presents itself as unambiguously negative (Doosje et al., 1998). A growing body of research suggests that learning about the links between past practices and present circumstances is a key factor in this. For example, German participants experienced more guilt and expressed greater willingness to make reparations to contemporary tribe members in today's Namibia when they had learned about the negative consequences of the genocide against the Nama and Herero (Imhoff et al., 2013).

Accordingly, a lack of critical historical knowledge and a motivation to protect group esteem predicted White people's denial of (systemic) present-day racism (Nelson et al., 2013). A recent experimental intervention found that this denial can be reduced by teaching White participants critical historical knowledge (Bonam et al., 2019, Study 2).

However, drawing from history in this way may have limited effects among people who are defensive against such reminders and favor historical closure, that is, to end a discussion about the past and its effects on the present (e.g., Kazarovytska & Imhoff, 2022). As such, people who favor historical closure might react defensively to being confronted with connections between the past and the present. The effectiveness of a contextualized confrontation might therefore be lower among people who favor such historical closure. Knowing the limitations of the contextualized confrontation will increase its applicability in practical settings.

Motivating moral redemption

Past research showed that the framing of morally threatening messages affects which strategies people choose to uphold their group's moral integrity (Ellemers et al., 2002). For example, people are more likely to engage in efforts for self-improvement and improvement of their social relations when moral failure is perceived to be repairable (Gausel & Leach, 2011; Gausel et al., 2012). Also, when equality is framed as a moral ideal (compared to moral obligation), Whites react with less identity threat and favor more cultural diversity as well as affirmative action (Does et al., 2011; Does et al., 2012). We believe that providing historical context to White privilege confrontations emphasizes the opportunity for individual moral redemption via political solidarity. Specifically, we assume that a contextualized confrontation can motivate collective action in solidarity with the disadvantaged, because it highlights the immorality of inequality (Agostini & van Zomeren, 2021), which should activate social justice morality motives (Janoff-Bulman & Carnes, 2013b; Janoff-Bulman et al., 2008). Social justice is a key motivation to engage in allyship and political solidarity (Goodman, 2011; Radke et al., 2020).

Social justice moral motives are rooted in a sense of communal responsibility, "activating group-based efforts to help with a particular focus on equality-oriented distributional justice" (Janoff-Bulman & Carnes, 2013b, p. 222). They have so far been conceptualized as rather stable characteristics, closely connected to political ideology. As such, social justice moral motives have not been tested as outcomes of confrontations with inequality nor privilege specifically. However, research indicates that preferences for

redistribution are highly sensitive to contextual information and background beliefs (Mitchell & Tetlock, 2009). A recent series of imagined-societies experiments showed that when people appraise an unequal distribution of resources to be based solely on luck, even self-identified "right-wingers" supported redistribution to a degree of "almost complete communism" (Nettle & Saxe, 2020, p. 11). Taken together, we hypothesize that a contextualized confrontation will lead to stronger endorsement of social justice morality and more acceptance of personal privilege compared to a control condition (H2a). Further, we expect that social justice morality will explain the positive effect of the contextualized confrontation on acceptance of personal privilege and solidarity (H2b).

The present research

Aims and hypotheses

In the present research, we argue that contextualizing White privilege in national colonial history is an effective intervention strategy that will have positive outcomes, whereas an intervention that only confronts with evidence of White privilege will trigger defensive reactions. We expect the contextualized confrontation to increase social justice morality, which in turn should positively affect privilege acceptance and lead to higher intentions to engage in political solidarity. In contrast, we expect a confrontation with mere evidence of privilege to have negative effects on these outcomes, explained by increased claims of life hardships. To compare, we test the new contextualized confrontation and the mere evidence confrontation against an active control condition.

Most of the work on White privilege stems from the US where racial discourse has been far more prevalent than in most European countries (Ball et al., 2022). This does not mean that European nations have overcome the issue of structural racism. Indeed, critical Whiteness scholars argue that Whiteness is an invisible but defining element of national and European identities (Arndt, 2020). This highlights the importance of research on White privilege interventions in European countries. We therefore test our ideas in Germany, where White privilege has been underexplored so far (Roig, 2016).

Overview of studies

Study 1 (N = 271, heterogenous German community sample) provides a conceptual replication of two experiments conducted in the US (Phillips & Lowery, 2015) and the UK (Murdoch & McAloney-Kocaman, 2019), which used the mere evidence confrontation intervention. Study 2 (N = 135, German student sample) conceptually replicates Study 1 and

additionally introduces the contextualized confrontation to investigate the role of social justice morality as potential mediator between the contextualized confrontation and acceptance of personal privilege and intentions to engage in solidarity-based collective action. In Study 3 (N = 347, heterogenous German community sample) we replicated the model of Study 1 in a demographically more diverse sample and considered more explicitly the moderating role of historical closure. Based on a large percentage of White German participants with a migration background in Study 3, and the likelihood that these participants have less (or a more critical) attachment to national identity, we also explored differences in how these groups respond to the confrontations.

The data of all three studies is available via an open science framework repository. Study 1 and 3 have also been preregistered (see Open Practice statement at the end of this work). All three studies adhere to the Declaration of Helsinki and participants gave their informed consent before the start of the study and were debriefed in written form after completing it.

Study 1

Responses to privilege confrontations seem to be motivated by identity protection (Knowles et al., 2014). White US-Americans who were confronted with mere evidence of White privilege reported higher levels of life hardships, were less likely to acknowledge personal privilege, and showed less support for pro-egalitarian policies (Philipps & Lowery, 2015). A similar study among White British people corroborated the finding "that exposure to evidence of White privilege is associated with higher levels of reported life hardships and lower beliefs in personal privilege" (Murdoch & McAloney-Kocaman, 2019, p. 208). One criticism of both previous studies is the lack of an active control condition (Strupp-Levitsky, 2020). Empty control conditions have been shown to lead to an overestimation of effects, simply because completing an exercise can increase engagement in a study (Lane et al., 2021). This is especially concerning as at least one previous study comparing a mere evidence confrontation with an active control condition has produced small or null-effects (Puryear et al., 2019). To address these limitations of previous work, we used an active control condition to our study design.

Method

Participants

Philipps and Lowery (2015) found an increase in life hardships after confrontation with evidence of White privilege using $N = 94$ (Experiment 1a) and $N = 91$ participants (Experiment 1b) in an online experiment. Murdoch and McAloney-Kocaman (2019), in their similar online experiment, had $N = 148$ participants. We preregistered to collect data from $N = 220$ participants (or 4 weeks of data collection) to ensure that, after exclusions, our sample size would be comparable to that of the latter study. We conducted a post-hoc power analysis with G*Power 3 (Faul et al., 2007) on the effect of manipulation on life hardships reported in both previous studies. This procedure indicated statistical power of $1-\beta = .63$ in the UK-based experiment (Murdoch & McAloney-Kocaman, 2019) as well as $1-\beta = .65$ (Experiment 1) and $1-\beta = .48$ (Experiment 2) in the US-based experiment (Phillips & Lowery, 2015) given the respective effect sizes and distribution of *N*. Based on the mean effect size of these three previous experiments ($d = .43$), power of Study 1 for this effect was expected at $1-\beta = .77$.

A heterogenous community sample was recruited via social network snowballing for an online survey resulting in $N = 272$ completed online questionnaires. Because we collected more than the preregistered number of cases and there were more exclusions than anticipated, we report the analysis with the full sample here and the results based on the preregistered cut-off ($N = 220$) in the supplementary material (section A). The results are fundamentally similar and there were no changes in significance in any of the results. Following our preregistered criteria, we excluded data from $n = 23$ participants because they were not German residents, and $n = 1$ who indicated they do not speak German well. Next, $n = 15$ were excluded because they did not self-identify as White. Another $n = 42$ were excluded who failed to give the correct answer in an attention check, finally $n = 18$ because they indicated they did not answer faithfully. There was no exclusion based on our time criterion (> 30 minutes for the entire study).

The final data consisted of $N = 164$ self-identified White participants (60.29% of the original sample), with a mean age of 43.28 ($SD = 15.96$) years. Our sample was thus roughly 10-15 years older than that of the two original studies. 62.2% were female, which is roughly 10% more than in the original studies. 20.1% of participants were students.

Design & procedure

Socio-demographic data was obtained in the beginning of the survey.[1] Next, participants were randomly assigned to one of two conditions in which different information was presented. The mere evidence condition was designed to be similar to a previous experiment conducted in the UK (Murdoch & McAloney-Kocaman, 2019). Here, participants were asked to read a short text about White privilege, followed by a list of seven statements about White privileges (adapted from McIntosh, 1988), such as "If I can't find affordable housing, I am certain that this is not due to my appearance or my last name". In the control condition, participants instead read a text about values that are shared among Germans, followed by a list of seven example values (see Frey, 2016), such as "I can freely choose my place of residence". Subsequently, all participants completed an attention check in which they were tasked to mark one out of four sentences that best represented what they had just read. Participants in the control condition who selected the item meant for the experimental condition were excluded to avoid unwanted effects of the attention check. Afterwards, we measured our dependent variables: life hardships, belief in White privilege, and acceptance of White privilege.

As a control question, at the end of the study we asked participants how much knowledge they had about the concept of White privilege before participating. Answers were given on a 4-point ordinal factor with "The term 'White privilege' was completely new to me", "I had heard about the term 'White privilege' but I did not know what it means, "I have already engaged a bit with White privilege prior to this study", and "I have already engaged extensively with White privilege prior to this study". The level of reported prior engagement in White privilege did not significantly differ between both conditions, $\chi^2(3) = 3.11$, $p = .37$. Afterwards, participants completed a sincerity check in which they were asked to mark one out of four statements that best represents how (in-)sincerely they had filled out this questionnaire. Finally, all participants were debriefed and dismissed.

Dependent variables

[1] We also measured social dominance orientation with the SDO-3 scale Pratto et al. (2013) on a range from 1 (does not apply at all) to 7 (applies completely). The scale had poor internal consistency ($\alpha = .45$) and was therefore dropped from further analysis.

All dependent variables were translated from the scales used in Phillips and Lowery (2015) and measured on a 7-point scale from 1 (does not apply at all) to 7 (applies completely). Measures were presented in the order they appear here.

Life hardships. We measured life hardships using five items (e.g., "My life has been full of hardships"; α = .85).

Belief in White privilege. We measured belief in White privilege with five items (e.g., "In our society, White people have certain advantages that minorities don't have"; α = .91).

Acceptance of White privilege. We used three items to tap into the acceptance of White privilege (e.g., "I have advantages in my life because I am White"; α = .94).

Results

Descriptive results together with correlations among the dependent variables appear in Table 1.

Table 1. Descriptive data of and correlations between dependent variables

		Condition	*N*	*M* (SD)	1	2
1	Life hardships	Confrontation	91	2.93 (1.29)	-	
		Control	71	2.92 (1.15)		
2	Acceptance of White Privilege	Confrontation	89	4.55 (1.82)	-.06	-
		Control	70	4.06 (2.09)		
3	Belief in White privilege	Confrontation	89	5.67 (1.27)	-.11	.66***
		Control	72	5.30 (1.50)		

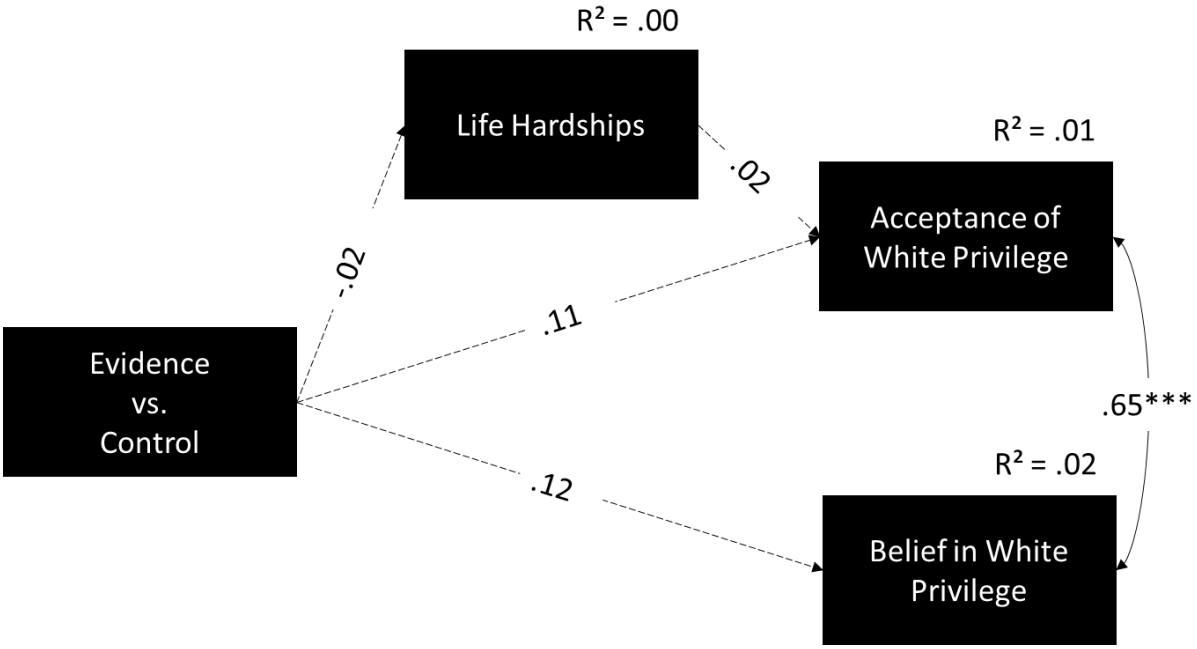
Note. *** $p < .001$.

We used jamovi v2.3 (The jamovi project, 2021)[2] in conjunction with the pathj module (Gallucci, 2021) to conduct a path analysis with maximum likelihood estimation and percentile bootstrapped confidence intervals based on $k = 5000$, which is the same setup used in the US-based publication (Phillips & Lowery, 2015, Experiment 2). Condition (1 = evidence confrontation, 0 = control) variable served as an exogenous factor. All dependent variables were mean-centered and regressed on condition. Acceptance of White privilege was also regressed on life hardships. The indirect effect was computed to test Hypothesis 1b. The model had acceptable fit statistics, $\chi^2(1) = 2.31$, $p = .13$; CFI = .99, RMSEA = .09, $p = .20$, 95% CI [.00 .25]; SRMR = .04.

[2] We preregistered the use of Mplus but opted for jamovi to keep all analyses in this paper within one statistical software. The procedure is otherwise unchanged.

Contrary to Hypothesis 1 and previous research, we did not find an effect of the mere evidence confrontation on life hardships, $b = 0.05$, $p = .78$, 95% CI [-0.42 0.33], $\beta = -.02$, nor on acceptance of White privilege, $b = 0.45$, $p = .16$, 95% CI [-0.20 1.08], $\beta = .11$. There was also no effect on belief in White privilege, $b = 0.34$, $p = .14$, 95% CI [-0.11 0.80], $\beta = .12$. Life hardships were unrelated to acceptance of White privilege, $b = 0.03$, $p = .76$, 95% CI [-0.16 0.23], $\beta = .02$. Accordingly, the indirect effect of condition on acceptance of White privilege through life hardships was not significant, $b = -0.002$, $p = .94$, 95% CI [-0.06 0.03], $\beta = -.00$. Figure 1 summarizes the path model.

Figure 1. Path model overview



Note. *** $p < .001$ Estimates are standardized. Dashed lines indicate non-significant paths.

To rule out that the lack of effects was due to the numerous deletions of participants who failed the attention check, we ran the model again without this exclusion criterion applied. In this analysis, the effects of the manipulation on the dependent variables was even smaller. Beyond that there was no difference to the main model reported above (see supplementary material, section A).

Discussion

Study 1 did not reproduce the effect specified in H1 that a mere evidence confrontation would increase life hardships and, through that, decrease acceptance of White privilege (Murdoch & McAloney-Kocaman, 2019; Phillips & Lowery, 2015). At least in this sample, life hardships were not indicative of a potential defensive reaction. One possible

explanation is that the existence of an active control condition reduced the effect of the mere evidence confrontation, which would mean that effects in previous research could be explained through higher engagement of participants in their experimental (compared to control) conditions.

Another possible explanation lies in the post-racial discourse in Germany. In Europe generally and in Germany specifically, there is a cultural tendency to understand "[…] racism as a phenomenon rooted in individual behavior as opposed to a systemic phenomenon shaped by structural factors" (Roig, 2016, p. 619). As such, participants may have appraised the White privilege confrontation as highlighting their individual level of racism (instead of unearned status) leading to life hardships being a less relevant (defensive) response. This potential explanation further highlights the benefit of contextualizing White privilege: If White privilege is not presented to Germans by merely providing them with evidence of its existence but by contextualizing it in Germany's involvement in colonialism, this might avoid appraisal of White privilege as an indicator for individual racism and instead foster moral engagement in social justice. We designed the following studies to examine more thoroughly the (potential) shift of moral motives in response to White privilege confrontations.

Study 2

In Study 2, we investigated whether the contextualized confrontation would indeed increase social justice morality but not individualized defensive reactions, such as claiming life hardships. To our knowledge, this is the first attempt to test the effectiveness of a contextualized confrontation strategy that embeds White privilege in colonialism. We also again examined the effect of a mere evidence confrontation to test the original hypothesis with another manipulation. Our main outcomes were acceptance of White privilege and solidarity-based collective action intentions. We expected that social justice morality would mediate a positive effect of the contextualized confrontation and that life hardships would mediate a negative effect of the mere evidence confrontation.

Method

Participants

A complete cohort of a first-year social psychology course at a German university was invited to participate in an online questionnaire during a lecture session and in exchange for course credit. $N = 183$ participated, of which $N = 148$ completed the survey. We removed data from n = 5 participants because they did not have German citizenship and $n = 8$ who self-

identified as People of Color. No case had to be removed based on a sincerity check (see below). The final sample consisted of N = 135 participants (91.3% of completed cases), who had a mean age of 21.64 (SD = 3.30) years. Most participants identified as female (n = 101, 76%) and one as "diverse/other" (1%).

Design & procedure

This study was part of a three-part online questionnaire with various social psychological research purposes. Demographic data, including self-identification as Black or Person of Color was collected at the beginning of the questionnaire. The part relevant to the current paper was the last part and introduced as being about attitudes on political and social issues.

Participants were randomly assigned to one of three conditions. In the contextualized confrontation condition, we presented information about Germany's involvement in colonial practices during the 20th century. We based the text on publicly available information provided by the German Federal Center for Political Education (Zimmerer, 2012). The text focused on how the exploitation of the Herero and Nama contributed to Germany's wealth. On a second page, participants received information about White privilege today including a short statement on its post-colonial context. To illustrate White privilege, we used a graphical depiction of White privilege adapted from Puryear et al. (2019). In the mere evidence condition, participants only received the information on White privilege without any mention of colonialism. In the control condition, participants read an information about Germany's voting rights reform in the 1970s, where the minimum age of being eligible to vote was lowered from 21 years to 18 years. This was followed on a second page by a short text about a current public debate to lower voting age to 16 years. This control condition was chosen to keep constant the historical and political context while reducing moral relevance and especially avoiding salience of racism. The manipulation materials can be reviewed in the supplementary material (section D). Following the dependent variables, a control question on prior engagement with Germany's history of colonialism was administered (on a scale from 1 = not at all to 7 = a lot). The mean in the contextualized confrontation condition (M = 4.11, SD = 1.29) was slightly higher than in the evidence confrontation (M = 3.72, SD = 1.41) and control condition (M = 3.67, SD = 1.44), but this difference was not significant, $F(2,132)$ = 1.36, p = 26, η^2_p = .02.

Dependent variables

Measures were presented in the order they appear here. Belief in White privilege was also measured (see supplementary material, section B) but dropped from analysis, because of its high correlation with acceptance of White privilege, $r(131) = .78$, $p < .001$.

Life hardships. Life hardships were measured similarly to Study 1 but shortened to three items: "My life has been full of hardships", "My life was easy" (reverse coded), "There were many difficulties in my life that I could not overcome" ($\alpha = .76$).

Social justice morality. We used three items from the Social Justice Morality subscale of the Moral Motives scales (Janoff-Bulmann et al., 2009): "We should all be responsible for improving the welfare of others beyond our immediate circle of friends and family", "It's an obligation, not a matter of personal preference, to provide for people worse off even if we're not close to them", and "It's important for those who are better off in society to work hard to provide more resources for those who are worse off" ($\alpha = .73$).

Acceptance of White privilege. Acceptance of White privilege was measured with the same three items as in Study 1. We reversed one item: "I have *not at all* profited from being White" ($\alpha = .88$).

Solidarity-based collective action intentions. Solidarity-based collective action intentions were measured with six items adapted from frequently used scales in the collective action literature (Becker et al., 2013). Participants were asked how likely it would be for them to participate in various forms of political activism in support of Black people in Germany if they had the chance to do so within the following months, e.g., "How likely is it that you will participate in a protest or demonstration" ($\alpha = .78$). Answers were measured on a 7-point scale ranging from -3 = very unlikely to 3 = very likely and recoded to values of 1 to 7.

Results

We conducted a MANOVA on all relevant outcomes. The value of Wilks' Lambda ($\Lambda = 0.89$) approached significance, $F(8,252) = 1.87$, $p = .065$, and the value of Roy's Largest Root (0.11) was significant, $F(4,127) = 3.61$, $p = .008$. As such there seemed to be a small multivariate effect on the main outcomes. We then computed univariate ANOVAs. Contrary to our hypothesis, there was no effect of condition on acceptance of White privilege nor on solidarity-based collective action intentions. Contrary to Hypothesis 1 there was also no effect of condition on life hardships $F(2,132) = 0.035$, $p = .70$, $\eta^2_p = .005$. However, in line with Hypothesis 2, the confrontations affected social justice morality differently, $F(2,132) = 6.42$, $p = .002$, $\eta^2_p = .09$. Tukey-corrected post-hoc tests revealed that social justice morality was

higher in the contextualized condition compared to the control condition, $t(132) = -3.57$, $p = .001$, $d = -0.76$, 95% CI [-1.19, -0.33]. The evidence confrontation did not affect social justice morality compared to the control condition, $t(132) = -1.59$, $p = .26$, $d = -0.34$, 95% CI [-0.76, 0.09]. There was no significant difference between both confrontations but Cohen's d indicated a medium-sized effect of higher social justice morality in the contextualized condition, $t(132) = 2.02$, $p = .11$, $d = 0.42$, 95% CI [0.01, 0.84]. Mean differences and ANOVA results are summarized in Table 1. Table 2 provides an overview of correlations between all dependent variables.

Table 1. Means, standard deviations, and results of univariate ANOVAs for relevant dependent variables.

	Condition	*N*	*M* (SD)	F (df1,df2)	*p*	η^2
Acceptance of White Privilege	Context	45	4.87 (1.46)	1.16 (2,129)	.316	.018
	Evidence	45	4.53 (1.89)			
	Control	42	4.33 (1.65)			
Solidarity-based Collective Action Intentions	Context	46	4.83 (1.22)	0.43 (2,132)	.653	.006
	Evidence	46	4.61 (1.29)			
	Control	43	4.62 (1.21)			
Life Hardships	Context	46	2.80 (1.14)	0.35 (2,132)	.704	.005
	Evidence	46	3.01 (1.42)			
	Control	43	2.87 (1.23)			
Social Justice Morality	Context	46	5.83 (0.89)	6.42 (2,132)	.002	.089
	Evidence	46	5.41 (0.99)			
	Control	43	5.07 (1.11)			

Table 2. Correlations between the relevant dependent variables

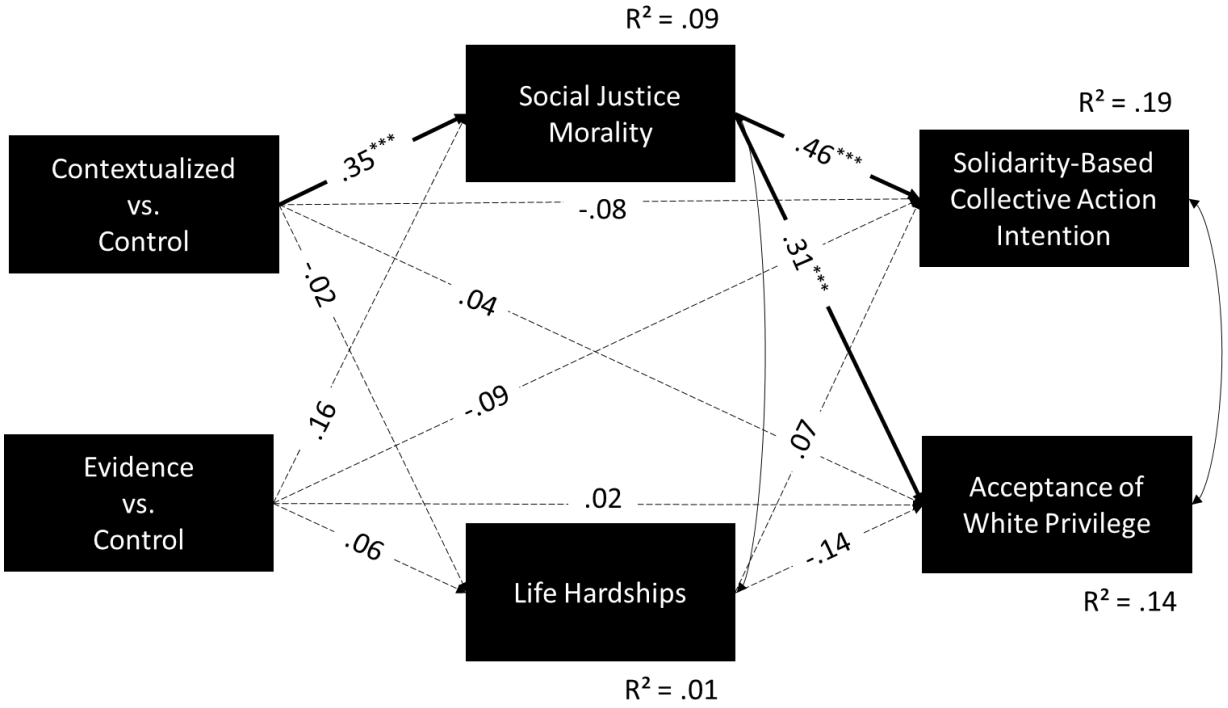
	1	2	3
1 Acceptance of White privilege	—		
2 Solidarity-Based Collective Action	.40***	—	
3 Life Hardships	-.19*	-.01	—
4 Social Justice Morality	.35***	.43***	-.18*

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

To investigate the mediating role of social justice morality and life hardships, we conducted a path analysis with robust maximum likelihood estimation. Condition was effect-

coded (intercept represents grand mean) with the control condition as reference category (Contextualized 1, Control -1; Evidence 1, Control -1). Social justice morality and life hardships were entered as parallel mediators between condition and acceptance of White privilege as well as solidarity-based collective action intentions. All dependent variables were mean-centered, and we specified a residual covariance between both mediators. The model had good model fit, $\chi^2(5) = 6.21$, $p = .29$, CFI = .98, SRMR = .04, RMSEA = .04, $p = .46$, 95% CI [.00, .13] and is depicted in Figure 2. A sensitivity power analysis was conducted for the indirect effect of the contextualized confrontation contrast through social justice morality on acceptance of White privilege. Based on a Monte Carlo power analysis (Schoemann et al., 2017) and the aggregated sample size of both conditions ($N = 89$), the analysis was powered for an indirect effect of $\beta_{ab} = .119$ under standard criteria (1- $\beta = .80$, $p < .05$).

Figure 2. Path model with social justice morality and life hardships as parallel mediators



Note. *** $p < .001$. Estimates are standardized. Dashed arrows represent non-significant paths.

Hypothesis 1b was not confirmed as life hardships did neither significantly affect acceptance of White privilege, $b = -0.18$, $p = .11$, 95% CI [-0.39 0.03], $\beta = -.14$, nor action intentions, $b = 0.07$, $p = .42$, 95% CI [-0.10 0.23], $\beta = .07$. The indirect effect via life

hardships was not significant. In line with Hypothesis 2b, social justice morality had a positive effect on both acceptance of White privilege $b = 0.51$, $p < .001$, 95% CI [0.27 0.77], $\beta = .32$, and intentions to engage in solidarity-based collective action, $b = 0.52$, $p < .001$, 95% CI [0.32 0.71], $\beta = .43$.

The indirect effects of the contextualized confrontation via social justice morality were significant for acceptance of White privilege, $b = 0.40$, $p = .005$, 95% CI [0.12 0.68], $\beta = .12$, as well as for solidarity-based collective action, $b = 0.40$, $p = .004$, 95% CI [0.13 0.68], $\beta = .15$. As expected, there was no significant indirect effect of the contextualized confrontation through life hardships. See Table 3 to review all indirect effects.

Table 3. Indirect effects through social justice morality and life hardships

	b	*p*	95% CI		β
Indirect effect of contextualized confrontation versus control					
SJM → Acceptance	0.40	< .01	0.12	0.69	.12
SJM → SBCA	0.40	< .01	0.13	0.68	.15
Hardships → Acceptance	0.01	.83	-0.08	0.10	.00
Hardships → SBCA	-0.00	.84	-0.04	0.03	-.00
Indirect effects of evidence confrontation versus control					
SJM → Acceptance	0.19	.12	-0.05	0.42	.05
SJM → SBCA	0.19	.12	-0.05	0.42	.07
Hardships → Acceptance	-0.03	.60	-0.14	0.08	-.01
Hardships → SBCA	0.01	.62	-0.03	0.05	.00

Note: IE = Indirect effects, SJM = Social justice morality, SBCA = Solidarity-based collective action

Discussion

Study 2 replicated and confirmed the results of Study 1 regarding the role of life hardships: Contrary to Hypothesis 1, there was no effect of the evidence confrontation on life hardships and life hardships were again not significantly related to acceptance of White privilege. Additionally, life hardships were also not significantly related to solidarity-based collective action intentions. Furthermore, supporting Hypothesis 2, we found an indirect effect of the contextualized confrontation on acceptance of White privilege and solidarity-based collective action intentions via social justice motives. However, there were no substantial total effects on these outcomes.

Study 2 relied on a student sample which might respond particularly positively to White privilege interventions. Therefore, it is worthwhile to check the generalizability of these findings among a more heterogeneous sample.

Study 3

We theorized that the contextualized confrontation should increase social justice motives, because it provides a systemic reasoning for White privilege. The student sample in Study 2 may have been especially prone to accept such systemic arguments (Brad et al., 2019). This is also indicated by a somewhat higher mean in privilege acceptance in Study 2 compared to Study 1. Recent research suggests that privilege confrontations often exacerbate but rarely change attitudes toward racial inequality (Puryear et al., 2019; Strupp-Levitsky, 2020). We therefore included additional variables to examine potential boundary conditions of the contextualized confrontation. One important moderator, for example, could be the attitude toward historical closure, because favoring historical closure should be associated with greater defensiveness towards confrontations with negative national history (Kazarovytska & Imhoff, 2022). We did not revise Hypothesis 1 despite the counter evidence from the two previous studies, because life hardships may still play a more important role as defensive reactions to White privilege in demographic groups that were underrepresented in our previous studies.

Method

Participants

We collected data from $N = 350$ participants online via Prolific. Inclusion criteria were self-reported White/Caucasian ethnicity and German nationality. Screening of the demographic data provided by Prolific revealed that $n = 2$ participants identified as Black or Mixed in a simplified ethnicity item. We excluded these cases and collected two additional participants to meet our preregistered sample of $N = 350$ White German adults. From this sample, we excluded data from $n = 3$ participants, because they failed one of two attention checks. $N = 4$ participants had a suspicious response speed (relative speed index above 2; see Leiner, 2019), but no conspicuous response patterns or open answers were identified. As such, in line with our preregistered criteria, we did not exclude these cases.

The final sample consisted of $N = 347$ participants with a mean age of 31.46 years ($SD = 9.99$, range: 18 to 67 years) of whom 35.3% identified as female. Political orientation (measured with a single item between 1 = left and 101 = right) was left leaning ($M = 36.31$, $SD = 19.78$) and subjective social status (Hoebel et al., 2015) was above the midpoint of a 1 = at the bottom to 10 = at the top scale ($M = 5.88$, $SD = 1.51$). There was an unexpected overrepresentation of 41% ($n = 141$) of participants who indicated to have a migration background (i.e., at least one parent did not have a German nationality). Because the

contextualized confrontation taps into German history, which may have different meaning for people with and without a history of migration, we additionally report results of the preregistered path analyses split between people with and without a migration background.

Design & procedure

To retain compatibility between Studies 1 and 2, we only slightly adapted the procedure: The text of the control condition was changed to account for recent political developments (the voting rights reform had passed without a reduction of minimum voting age). We also omitted the belief in White privilege measure because of its high correlation with acceptance of White privilege in Study 2.

Dependent variables

Measures were presented in the order they appear here. Response scales for all dependent variables ranged between 1 (disagree completely) and 7 (agree completely). Life hardships ($\alpha = .85$), social justice morality ($\alpha = .80$), acceptance of White privilege ($\alpha = .93$), and solidarity-based collective action intention ($\alpha = .85$) were measured exactly as in Study 1. We additionally measured allyship, anger at injustice and White guilt, text-based responses on the reasoning for racial inequality, and White responsibility after the main dependent variables. More information on these measures can be found in the supplementary materials (section C).

Additional variables

The following variables were measured after the dependent variables.

Preference for historical closure. To check for a potential moderation effect, we measured preference for historical closure with the seven-items Future Orientation subscale of the Historical Closure scale developed by Kazarovytska and Imhoff (2022). Example item: “It is better to talk about events from the present than events from the past” ($\alpha = .95$).

Control questions. As additional control questions, we asked participants (after administering the dependent variables) how much they had engaged with each theme presented in the three conditions prior to their study participation: German voting rights, White privilege, and Germany’s history of colonialism. There was no significant difference between the three conditions concerning prior engagement in German voting rights, $F(2,344) = 2.69$, $p = .34$, $\eta^2 = .01$, White privilege, $F(2,344) = 2.86$, $p = .52$, $\eta^2 = .00$, and Germany’s history of colonialism, $F(2,344) = 2.71$, $p = .40$, $\eta^2 = .01$.

Results

Effects of condition on dependent variables

We conducted a MANOVA to test the overall effect of condition on the relevant dependent variables. Both Wilks' lambda, $\Lambda = 0.95$, $F(8,680) = 2.25$, $p = .023$, and Roy's Largest Root, value = 0.05, $F(4,341) = 4.29$, $p = .002$, were significant indicating a multivariate effect of our manipulation. Contrary to our predictions, however, post-hoc ANOVAs revealed no univariate effects of the conditions regarding acceptance of White privilege nor solidarity-based collective action intentions (preregistered analysis). There was also no significant mean difference on the other main dependent variables (see Table 4). Correlations between the relevant dependent variables appear in Table 5.

Table 4. Descriptive information and ANOVA results on relevant dependent variables

	Condition	*N*	*M* (SD)	ANOVA		
				F (df1,df2)	*p*	η^2
Acceptance of White Privilege	Context	117	4.09 (1.87)	1.65 (2,343)	.19	.010
	Evidence	112	4.52 (1.77)			
	Control	117	4.36 (1.86)			
Solidarity-Based Collective Action Intentions	Context	117	4.21 (1.57)	1.84 (2,344)	.16	.011
	Evidence	112	3.98 (1.49)			
	Control	118	3.84 (1.38)			
Life Hardships	Context	117	3.48 (1.31)	1.31 (2,344)	.27	.008
	Evidence	112	3.48 (1.31)			
	Control	118	3.77 (1.41)			
Social Justice Morality	Context	117	5.26 (1.24)	1.20 (2,344)	.30	.007
	Evidence	112	5.11 (1.14)			
	Control	118	5.03 (1.06)			

Table 5. Correlations between the main dependent variables

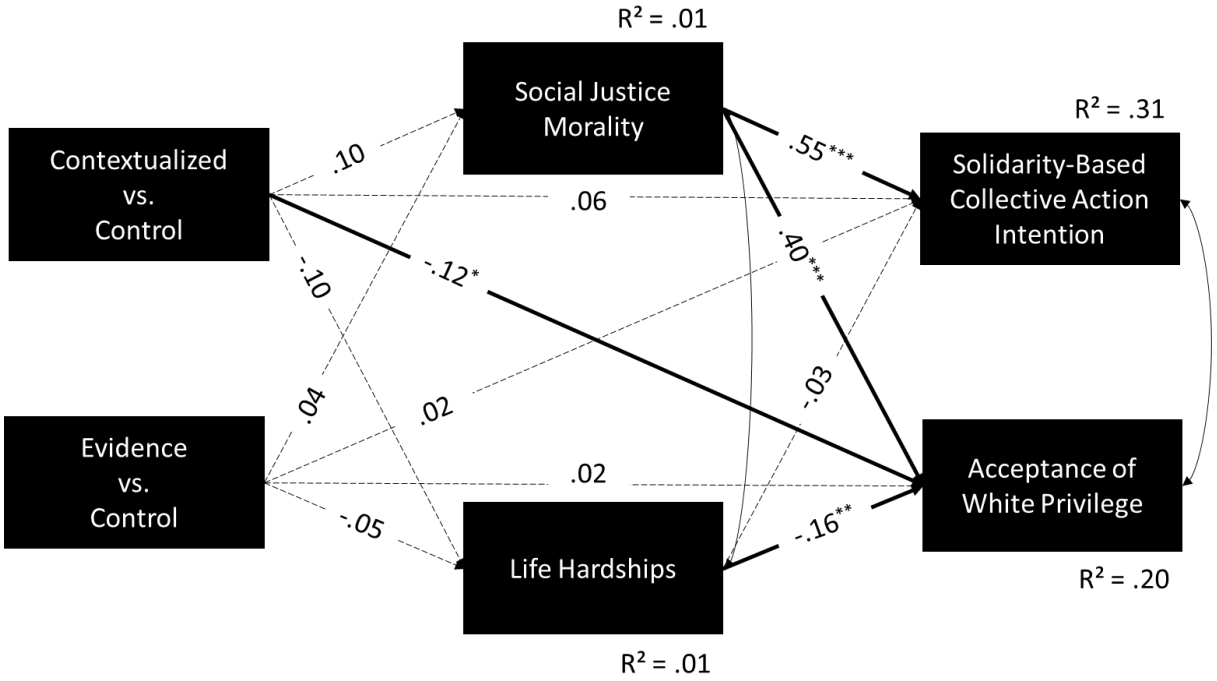
	1	2	3
1 Privilege Acceptance	-		
2 Solidarity-Based Collective Action Intentions	.49***	-	
3 Hardships	-.16**	-.05	-
4 Social Justice Morality	.39***	.55***	-.03

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

Path analyses

As preregistered, we used the same path analysis procedure as in Study 2 to investigate the mediating role of life hardships and social justice morality. The effects of condition in this preregistered model were small and mostly non-significant but there was an unexpected negative path of the contextualized condition on acceptance of White privilege. Like in Study 2, social justice morality was positively related to acceptance of White privilege and solidarity-based collective action intentions. Additionally, we found support for the negative association between life hardships and acceptance of White privilege. The model and results are summarized in Figure 3. All indirect effects were non-significant ($p > .05$, see supplementary material, section C).

Figure 3. Preregistered path model



Note. * $p < .05$, *** $p < .001$. Estimates are standardized.

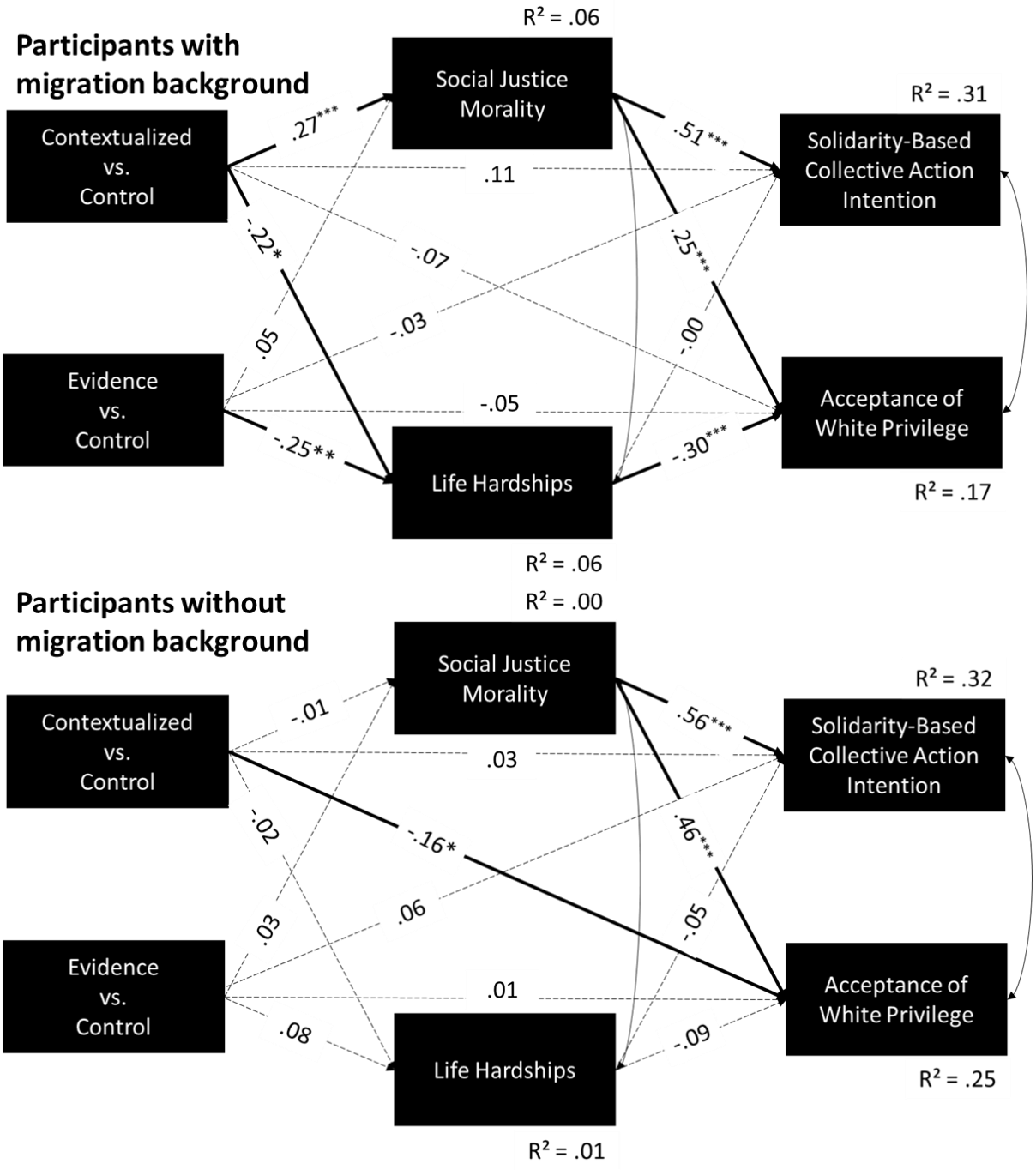
One possible explanation of the lack of effects lies in the overrepresentation of people with a migration background, who might react more openly to the confrontation with negative national history. We therefore repeated the above analysis but split participants into groups with ($N = 141$) and without ($N = 206$) an immediate family history of migration (not preregistered). In this multiple-group path analysis we allowed all estimates to vary freely between groups. Results within White Germans without a migration background were similar

to the overall model and, as such, mostly non-significant. However, White Germans with a migration background reacted more strongly to the confrontations: There was a significant positive indirect effect of the contextualized confrontation on solidarity-based collective action intention through social justice morality, $b = 0.42$, $p = .01$, 95% CI [0.10, 0.75], $\beta = .14$, and a marginal indirect effect on privilege acceptance, $b = 0.25$, $p = .051$, 95% CI [-0.00 0.50], $\beta = .07$. Surprisingly, there was also a *negative* effect of both conditions on life hardships, which translated to a positive indirect effect on privilege acceptance of both the contextualized confrontation, $b = 0.24$, $p = .045$, 95% CI [0.01 0.48], $\beta = .07$, and of the mere evidence confrontation, $b = 0.29$, $p = .02$, 95% CI [0.05 0.54], $\beta = .08$. All other indirect paths were non-significant (see supplementary material section C for details). Taken together, within White Germans with a migration background we found the opposite pattern of Hypothesis 1 and results that align with Hypothesis 2. The results are summarized in Figure 4.

To rule out alternative explanations, we ensured that participants with and without migration background were equally distributed among the three conditions, $\chi^2(2) = 1.22$, $p = .55$. There were also no overall mean differences between participants with and without migration background on our main dependent variables: privilege acceptance, $t(307.20) = 0.09$, $p = .93$, $d = .01$, solidarity-based collective action intentions, $t(296.93) = -0.06$, $p = .97$, $d = -.006$, life hardships, $t(302.71) = -0.06$, $p = .95$, $d = -0.007$, and social justice morality, $t(323.81) = 0.76$, $p = .45$, $d = .08$.[3]

Figure 4. Path analyses split for White Germans with migration background ($N = 141$) and without ($N = 206$)

[3] We also checked whether migration background was associated with demographic and other third variables that might contribute to a difference between both groups. T-Tests between participants with and without migration background revealed no significant differences on political orientation, social-economic status, historical closure, and on the three control variables. However, participants with a migration background were significantly younger, $t(344.23) = -3.96$, $p < .001$, $d = -0.42$. Still, the pattern of results of the split path models was similar when controlling for age (changes pertained only to the third decimal, no changes in [in-]significance of effects), indicating that migration background explains variance over and above age differences.



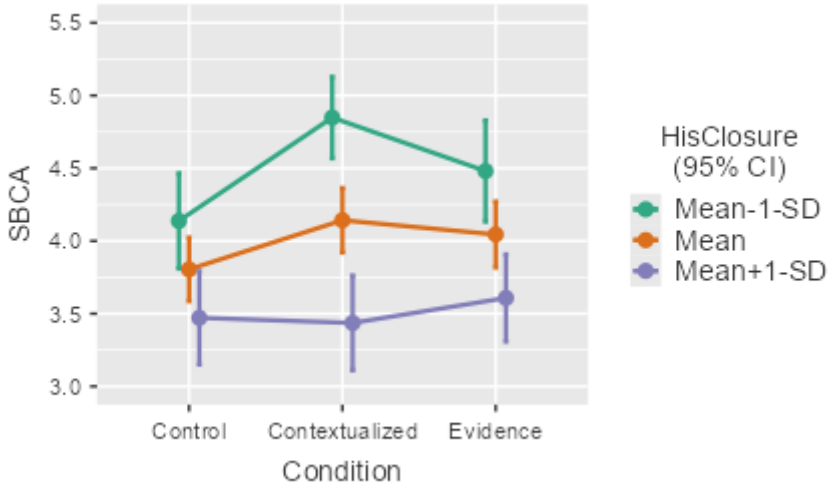
Note. * $p < .05$, *** $p < .001$. Estimates are standardized.

Examining the moderating role of historical closure.

We preregistered an investigation into whether historical closure would interact with the effect of our contextualized manipulation on social justice morality. A general linear model with the condition factor (effect-coded), historical closure (z-standardized), and their interaction revealed no significant interaction effect, neither of the contextualized versus control effect, b = -0.14, p = .13, 95% CI [-0.33, 0.04], β = -.18, nor of the mere evidence versus control effect, b = -0.02, p = .88, 95% CI [-0.21, 0.18], β = -.02. There was also no

main effect of condition, $F(2,341) = 0.63$, $p = .53$, $\eta^2_p = .004$. However, historical closure was positively related to social justice morality, $F(1,341) = 74.87$, $p < .001$, $\eta^2_p = .18$. Due to the significant negative effect of the contextualized condition on solidarity-based collective action intention in the overall path model, we repeated this analysis with solidarity-based collective action intentions as dependent variable (not preregistered). This analysis found no significant effect of condition, $F(2,341) = 1.11$, $p = .33$, $\eta^2_p = .006$, a significant effect of historical closure, $F(1,341) = 80.79$, $p < .001$, $\eta^2_p = .19$, and a significant interaction effect between condition and historical closure, $F(2,341) = 3.05$, $p = .049$, $\eta^2_p = .018$. The interaction was significant for the contextualized condition versus control condition dummy, $b = -.29$, $p = .02$, 95% CI [-0.52, -0.05], $\beta = -.28$. As can be seen in Figure 5, the negative correlation between historical closure and solidarity-based collective action intentions was higher in the contextualized condition.

Figure 5. Interaction effect of condition and historical closure on solidarity-based collective action



Note. HisClosure = Historical closure, SBCA = solidarity-based collective action intentions

Discussion

The results of Study 3 were mixed regarding our hypotheses. While a multivariate analysis revealed that the pattern of a linear composite of social justice morality, life hardships, acceptance of White privilege, and solidarity-based collective action intentions varied between the three conditions, we did not observe total univariate effects on these outcomes. The path analysis, as in both previous experiments, did not provide support for Hypothesis 1: The mere evidence confrontation did not affect claims of life hardship. Contrary to Hypothesis 2 and Study 2, there was no effect of the contextualized confrontation

on social justice morality. We also found an unexpected negative direct effect of the contextualized confrontation on acceptance of White privilege in the path model.

Some additional analyses shed more light on these results: A multiple group path analysis revealed that the results differed strongly between White Germans with and without a migration background. First, the overall pattern of results largely reflected the pattern found within White Germans without a migration background. Second, among White Germans with a migration background, but not among those without migration background, both confrontations *reduced* claims of life hardships and life hardships were associated with less acceptance of White privilege. Third, in the migration background group, we found the hypothesized positive effect of the contextualized confrontation through social justice morality. Additionally, there was a significant interaction between the experimental condition and historical closure indicating that the contextualized confrontation had a more positive effect on solidarity-based collective action intentions among participants who tended to reject historical closure arguments.

Taken together, Study 3 provides additional evidence that White Germans do *not* react with life hardships to privilege confrontations. In fact, White Germans with a migration background tended to report even less life hardships after a confrontation. This could be reflective of a less individualized understanding of their social position. Study 3 also provides some important boundary conditions of the contextualized confrontation: Overall, the contextualized confrontation did not produce positive effects. However, a closer look at the interaction with participant demographics revealed that the contextualized confrontation had the expected positive effects within participants with a migration background and who are more supportive of a critical reflection of Germany's history. Accordingly, it seems that the positive effects of contextualizing White privilege in national colonial history depend on White individuals' willingness to critically engage with their nation's past.

General Discussion

In the current paper, we examined the effect of two different strategies of confronting White people with White privilege on their level of privilege acceptance and political solidarity: Presenting mere evidence of White privilege and contextualizing White privilege in its colonial roots. Whereas the mere evidence confrontation has received considerable attention in past research (Murdoch & McAloney-Kocaman, 2019; Phillips & Lowery, 2015), studies have so far not examined the effectiveness of communicating White privilege in a post-colonial framework. Our main aims were therefore to investigate a) the defensive effects

of a mere evidence confrontation further and b) the potential constructive outcomes of a contextualized strategy via social justice morality.

Across all three experiments we did not find support that a mere evidence confrontation leads to claims of life hardships and therefore less acceptance of White privilege (Hypothesis 1) as suggested by previous work (Murdoch & McAloney-Kocaman, 2019; Phillips & Lowery, 2015), even though our experiments had comparable or higher power than the original studies to find this effect. Additionally, life hardships were unrelated to acceptance of White privilege in both Study 1 and Study 2. We found this negative association only in Study 3 and only among participants with a migration background. Interestingly, in this group both confrontations decreased claims of life hardships compared to a control condition. All in all, White Germans do not seem inclined to distance themselves from White privilege via claiming life hardships.

Regarding our new contextualized confrontation, we found some positive effects and several boundary conditions. Among a student sample (Study 2), the contextualized confrontation increased social justice morality, which was associated with higher acceptance of White privilege and intentions to engage in solidarity-based collective action as expected in Hypothesis 2. However, this effect did not replicate among the community sample in Study 3. In fact, there was a negative direct effect of the contextualized confrontation on solidarity-based collective action intentions. Further analyses revealed that the positive effect of the contextualized confrontation was limited to White participants with a migration background and that White participants who were more open to critically reflect Germany's past were more likely to report solidarity-based collective action intentions after the contextualized confrontation.

On the role of meritocratic defense in White Germans

We did not find that White privilege confrontations increase claims of life hardships among White Germans. In previous research, self-affirmation prior to a privilege confrontation had reduced the defensive effect of the mere evidence confrontation. Therefore, claiming life hardships has been interpreted as an attempt to preserve self-esteem (Phillips & Lowery, 2015, 2020). It could be that in the German context, albeit a meritocratic culture as well, justifying one's status might not be the most prominent strategy to defend self-esteem. Given the high salience of Germany's national-socialist history and the strong cultural avoidance of racial discourse in continental Europe (Ball et al., 2022; Roig, 2016), moral concerns might prevent White Germans from reacting on the level of status justification to

White privilege confrontations. Another explanation might be that in Germany meritocracy is linked to "work ethic", whereas in the UK meritocracy is perceived as a "status ethic" (Heuer et al., 2020). Specifically, British people seem to rely more strongly on the size of wages ("if you perform better than the other one, you get more money") whereas Germans highlight the importance of work ("work is the elixir of life"). As such, the meritocratic defense among White Germans may rather be found when investigating their support for a protestant work ethic after a confrontation with their privilege. In any case, the role of meritocratic threat in the 3D model seems to depend, contrary to its claims, on cultural context even within "Western" nations.

The bounded effectiveness of contextualizing White privilege

Overall, the positive effect of the contextualized confrontation was less stable than expected. Our experiments indicate that the effectiveness of the confrontation depends strongly on the relation and attitude of the confronted individual toward their nation's history. This was indicated by a stronger effect of the confrontation among White German students, among White Germans with a migration background, and among White Germans who do not support historical closure. It could be the case that White Germans who grew up with at least one non-German parent tend to have a lower and/or more critical identification with Germany. However, research on White migrants is sparse because in many Western countries, migration is often associated in terms of ethnic or racial differences (Lundström, 2017). Still, one could speculate that a more critical stance toward one's national identity increased the likelihood of accepting White privilege and engaging in political solidarity. At least for White German students and for White Germans who reject historical closure, this might be motivated via the opportunity for moral redemption. In a more general sense, these boundaries of the contextualized confrontation demonstrate the importance of keeping negative collective memories alive to preserve the collective society's motive to attain social justice (Licata et al., 2018; Liu & Hilton, 2005) and to counteract the ideological and structural self-justification of systems of inequality (Täuber & Moughalian, 2022).

Future research

Our results resonate with and extend previous research that found that the effects of privilege interventions strongly depend on previously held attitudes toward racial inequality (Brad et al., 2019; Puryear et al., 2019; Strupp-Levitsky, 2020). We showed that only if the specific style of the confrontation (i.e., including a connection to negative national history) matches the participants' attitudes (i.e., openness toward critical reflection of collective

memory), the confrontation produced positive effects. Future research should therefore follow-up on the present findings by examining how tailoring privilege confrontations to their audience can increase their effectiveness.

One important argument made in the 3D model is that White privilege is *not* invisible but actively avoided by White individuals (Knowles et al., 2014). Arguably, themes surrounding *race* and racism are even more strongly avoided in Germany than in the US, and racism also seems to be even more invisible in Germany specifically but also in Europe generally (Roig, 2016; Ball et al., 2022). This invisibility is maintained through a systemic exclusion of colonialism from collective memory (Krüger, 2004. For example, colonialism has for a long time been circumvented and misrepresented in German school textbooks (Marmer et al., 2010); so-called "colonized curricula" (Weiner, 2016). The collective avoidance of a discourse on racism and the racialization of German (and European) identity as White has remained unchallenged to this day (Arndt, 2020; Nedelsky, 2023). This highlights the importance to avoid overgeneralization and essentialization of psychological effects. Instead, research should adopt the perspective that social psychological processes are bound to the specific systems they are observed in (Leach & Bou Zeineddine, 2021). As almost all social psychological research on White privilege stems from the US, we advocate for more research on how cultural attitudes toward racial inequality shape responses to White privilege across "Western" contexts and beyond. A research synthesis project that collects published and unpublished data on attitudes toward White privilege across different context would for example provide important first insights into the real-world effects of confrontations with White privilege and racial injustice.

Limitations

We tested our hypotheses across different populations and with bigger sample sizes than those that had been used in similar experiments, which provides some confidence in the lack of significant effects of the mere evidence confrontation on life hardships across all three experiments. However, some limitations must be addressed, specifically regarding the role of causality and unexplored mechanisms.

We have relied on path analyses as means to investigate mediation hypotheses in all three experiments. Although this approach is viable to detect correlational patterns that indicate the presence of mediation, it can neither confirm the existence of causal mediation nor the model's correct specification (Fiedler et al., 2011). The indirect effect of the contextualized confrontation on solidarity-based collective action intentions that we found in

some samples is only one of many possible causal explanations, because the test of association between dependent variable and mediator is correlational. To ensure that social justice morality is in fact a relevant mediator, future research should ideally manipulate this mediator in conjunction with the independent variable in a so-called parallel design (Danner et al., 2015; Pirlott & MacKinnon, 2016).

Relatedly, one can argue that the contextualized confrontation was a rather indirect strategy of activating social justice morality. We assumed that contextualizing White privilege in colonialism emphasizes collective rather than individual responsibility for White privilege and racial inequality, which would allow one to challenge the immorality of privilege by acting against the systems that perpetuate it. However, we this only indirectly via social justice morality motives, which is a collective dimension of morality (Janoff-Bulman & Carnes, 2013a, 2013b). Future research might consider more direct ways to measure the perception of collective versus individual responsibility in response to such confrontations.

Previous research on morality also indicated that fear of rejection is an important driver for defensive responses to moral failure (Gausel et al., 2012). Therefore, it would be interesting to see how appraisals differ between confrontations that provide an opportunity for moral redemption, for example by framing political action as an opportunity to self-enhance, and those that do not provide such an opportunity (essentially extending the design used in Does et al., 2011). Lastly, we did not directly test our assumption that the positive effects of the contextualized confrontation were limited specifically to those who show a more critical identification with their nation/group. This is a post-hoc reasoning based on different results among different demographic groups and people with more openness to engage with negative national history. Future research should attempt to identify whether such critical attachment to national identity truly increases the effectiveness of a contextualized White privilege confrontation.

Conclusion

In this paper, we investigated the effects of two different strategies of White privilege confrontations in a German context: a confrontation with evidence of White privilege as it has been used in previous research (Murdoch & McAloney-Kocaman, 2019; Phillips & Lowery, 2015) and a confrontation that contextualizes White privilege in Germany's colonial past. Results across three experiments indicate that Germans are not inclined to distance from White privilege by claiming life hardships. Life hardships were also only inconsistently related to acceptance of White privilege. Furthermore, we examined whether contextualizing

White privilege in its historical origins triggers social justice morality, which we expected to positively affect the level of White privilege acceptance and political solidarity. We found support for this hypothesis in White Germans that are likely to be open to receiving critical information about their national identity, such as students, Germans with a migration background, or those who do not support historical closure. Although the positive effects of the contextualized confrontation were limited by these boundary conditions, we emphasize that White privilege interventions *can* leverage negative collective history (such as the colonial history of many European nations) to moralize racial inequality, which is an important new pathway for future research and practice to test and adopt.

Open Practice

Data, figures, supplementary materials, and preregistrations are available at https://osf.io/dgnx4/?view_only=e47517aaa93b4c69a25c62afbd9853ad

References

Abrams, D., & Hogg, M. A. (1990). An introduction to the social identity approach. In D. Abrams & M. A. Hogg (Eds.), *Social identity theory: Constructive and critical advances.* Harvester Wheatsheaf.

Agostini, M., & van Zomeren, M. (2021). Toward a comprehensive and potentially cross-cultural model of why people engage in collective action: A quantitative research synthesis of four motivations and structural constraints. *Psychological Bulletin*, *147*(7), 667–700. https://doi.org/10.1037/bul0000256

Allpress, J. A., Brown, R., Giner-Sorolla, R., Deonna, J. A., & Teroni, F. (2014). Two faces of group-based shame: Moral shame and image shame differentially predict positive and negative orientations to ingroup wrongdoing. *Personality & Social Psychology Bulletin*, *40*(10), 1270–1284. https://doi.org/10.1177/0146167214540724

Arndt, S. (2020). Weißsein: Die verkannte Strukturkategorie Europas und Deutschlands. In M. M. Eggers, G. Kilomba, P. Piesche, & S. Arndt (Eds.), *Mythen, Masken und Subjekte: Kritische Weißseinsforschung in Deutschland* (4th ed., pp. 24–28). Unrast.

Ball, E., Steffens, M. C., & Niedlich, C. (2022). Racism in Europe: Characteristics and intersections with other social categories. *Frontiers in Psychology*, *13*, 789661. https://doi.org/10.3389/fpsyg.2022.789661

Becker, J. C., Wright, S. C., Lubensky, M. E., & Zhou, S. (2013). Friend or ally: Whether cross-group contact undermines collective action depends on what advantaged group members say (or don't say). *Personality & Social Psychology Bulletin*, *39*(4), 442–455. https://doi.org/10.1177/0146167213477155

Bergkamp, J., Martin, A., & Olson, L. (2022). Social privilege: Flipping the coin of inequity. In J. L. Chin (Ed.), *Race and Ethnicity in Psychology Ser. The Psychology of Inequity*. ABC-CLIO LLC.

Bergkamp, J., Olson, L., & Martin, A. (2022). Before allyship: A model of integrating awareness of a privileged social identity. *Frontiers in Psychology*, *13*, 993610. https://doi.org/10.3389/fpsyg.2022.993610

Bonam, C. M., Nair Das, V., Coleman, B. R., & Salter, P. (2019). Ignoring history, denying racism: Mounting evidence for the Marley hypothesis and epistemologies of ignorance. *Social Psychological and Personality Science*, *10*(2), 257–265. https://doi.org/10.1177/1948550617751583

Bonds, A., & Inwood, J. (2016). Beyond white privilege: Geographies of white supremacy and settler colonialism. *Progress in Human Geography*, *40*(6), 715–733. https://doi.org/10.1177/0309132515613166

Brad, L. C. E., Spisz, T. J., & Tanega, C. G. (2019). Does "privilege checking" make us less racist, or more? Generation and political orientation matter. *Race and Social Problems*, *11*(1), 1–14. https://doi.org/10.1007/s12552-018-9246-0

Case, K. A., Hensley, R., & Anderson, A. (2014). Reflecting on heterosexual and male privilege: Interventions to raise awareness. *Journal of Social Issues*, *70*(4), 722–740. https://doi.org/10.1111/josi.12088

Case, K. A., Iuzzini, J., & Hopkins, M. (2012). Systems of privilege: Intersections, awareness, and applications. *Journal of Social Issues*, *68*(1), 1–10. https://doi.org/10.1111/j.1540-4560.2011.01732.x

Danner, D., Hagemann, D., & Fiedler, K. (2015). Mediation analysis with structural equation models: Combining theory, design, and statistics. *European Journal of Social Psychology*, *45*(4), 460–481. https://doi.org/10.1002/ejsp.2106

Does, S., Derks, B., & Ellemers, N. (2011). Thou shalt not discriminate: How emphasizing moral ideals rather than obligations increases Whites' support for social equality. *Journal of Experimental Social Psychology*, *47*(3), 562–571. https://doi.org/10.1016/j.jesp.2010.12.024

Does, S., Derks, B., Ellemers, N., & Scheepers, D. (2012). At the heart of egalitarianism. *Social Psychological and Personality Science*, *3*(6), 747–753. https://doi.org/10.1177/1948550612438924

Doosje, B., Branscombe, N. R., Spears, R., & Manstead, A. S. R. (1998). Guilty by association: When one's group has a negative history. *Journal of Personality and Social Psychology*, *75*(4), 872–886. https://doi.org/10.1037/0022-3514.75.4.872

Eckerle, F., Rothers, A., Kutlaca, M [Maja], Henss, L., Agunyego, W., & Cohrs, J. C. (2023). Appraisal of male privilege: On the dual role of identity threat and shame in response to confrontations with male privilege. *Journal of Experimental Social Psychology*, *108*, 104492. https://doi.org/10.1016/j.jesp.2023.104492

Ehrke, F., Ashoee, A., Steffens, M. C., & Louvet, E. (2020). A brief diversity training: Raising awareness of ingroup privilege to improve attitudes towards disadvantaged outgroups. *International Journal of Psychology*, *55*(5), 732–742. https://doi.org/10.1002/ijop.12665

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology*, *53*, 161–186.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, *47*(6), 1231–1236. https://doi.org/10.1016/j.jesp.2011.05.007

Frey, D. (2016). *Psychologie der Werte*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-48014-4

Gallucci, M. (2021). *PATHj: jamovi Path Analysis* [Computer software]. https://pathj.github.io

Gausel, N., & Leach, C. W. (2011). Concern for self-image and social image in the management of moral failure: Rethinking shame. *European Journal of Social Psychology*, *41*(4), 468–478. https://doi.org/10.1002/ejsp.803

Gausel, N., Leach, C. W., Vignoles, V. L., & Brown, R. (2012). Defend or repair? Explaining responses to in-group moral failure by disentangling feelings of shame, rejection, and inferiority. *Journal of Personality and Social Psychology*, *102*(5), 941–960. https://doi.org/10.1037/a0027233

Goodman, D. J. (2011). *Promoting diversity and social justice: Educating people from privileged groups* (Second edition). Routledge. http://gbv.eblib.com/patron/FullRecord.aspx?p=683962

Heuer, J.-O., Lux, T., Mau, S., & Zimmermann, K. (2020). Legitimizing inequality. *Comparative Sociology*, *19*(4-5), 542–584. https://doi.org/10.1163/15691330-BJA10017

Hoebel, J., Müters, S., Kuntz, B., Lange, C., & Lampert, T. (2015). Messung des subjektiven sozialen Status in der Gesundheitsforschung mit einer deutschen Version der MacArthur Scale [Measuring subjective social status in health research with a German version of the MacArthur Scale]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, *58*(7), 749–757. https://doi.org/10.1007/s00103-015-2166-x

Imhoff, R., Wohl, M. J. A., & Erb, H.-P. (2013). When the past is far from dead: How ongoing consequences of genocides committed by the ingroup impact collective guilt. *Journal of Social Issues*, *69*(1), 74–91. https://doi.org/10.1111/josi.12004

The jamovi project. (2021). *jamovi* (Version 2.2) [Computer software]. https://www.jamovi.org

Janoff-Bulman, R., & Carnes, N. C. (2013a). Moral context matters: A reply to Graham. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *17*(3), 242–247. https://doi.org/10.1177/1088868313492021

Janoff-Bulman, R., & Carnes, N. C. (2013b). Surveying the moral landscape: Moral motives and group-based moralities. *Personality and Social Psychology Review*, *17*(3), 219–236. https://doi.org/10.1177/1088868313480274

Janoff-Bulman, R., Sheikh, S., & Baldacci, K. G. (2008). Mapping moral motives: Approach, avoidance, and political orientation. *Journal of Experimental Social Psychology*, *44*(4), 1091–1099. https://doi.org/10.1016/j.jesp.2007.11.003

Kazarovytska, F., & Imhoff, R. (2022). Three fish at one hook? Future-oriented, reconciliatory, and defensive claims for historical closure as expressions of the same defensive desire. *Personality & Social Psychology Bulletin*, 1461672221124674. https://doi.org/10.1177/01461672221124674

Knowles, E. D., Lowery, B. S., Chow, R. M., & Unzueta, M. M. (2014). Deny, distance, or dismantle? How White Americans manage a privileged identity. *Perspectives on Psychological Science*, *9*(6), 594–609. https://doi.org/10.1177/1745691614554658

Krüger, G. (2004). Vergessene Kriege: Warum gingen die deutschen Kolonialkriege nicht in das historische Gedächtnis der Deutschen ein? In D. Langewiesche (Ed.), Der Krieg in den Gründungsmythen europäischer Nationen und der USA (1st ed., 120-137). Campus Verlag.

Lane, A. M., Beedie, C. J., Devonport, T. J., & Friesen, A. P. (2021). Considerations of control groups: Comparing active-control with no treatment for examining the effects of brief intervention. *Sports*, *9*(11). https://doi.org/10.3390/sports9110156

Leach, C. W., & Bou Zeineddine, F. (2021). A systems view of emotion in socio-political context. *Affective Science*, *2*(4), 353–362. https://doi.org/10.1007/s42761-021-00051-z

Leach, C. W., Iyer, A., & Pedersen, A. (2006). Anger and guilt about ingroup advantage explain the willingness for political action. *Personality & Social Psychology Bulletin*, *32*(9), 1232–1245. https://doi.org/10.1177/0146167206289729

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in Internet surveys. *Survey Research Methods*, *13*(3), 229–248. https://doi.org/10.18148/SRM/2019.V13I3.7403

Licata, L., Khan, S. S., Lastrego, S., Cabecinhas, R., Valentim, J. P., & Liu, J. H. (2018). Social representations of colonialism in Africa and in Europe: Structure and relevance for contemporary intergroup relations. *International Journal of Intercultural Relations*, *62*, 68–79. https://doi.org/10.1016/j.ijintrel.2017.05.004

Liu, J. H., & Hilton, D. J. (2005). How the past weighs on the present: Social representations of history and their role in identity politics. *The British Journal of Social Psychology*, *44*(Pt 4), 537–556. https://doi.org/10.1348/014466605X27162

Lowery, B. S., Chow, R. M., Knowles, E. D., & Unzueta, M. M. (2012). Paying for positive group esteem: How inequity frames affect whites' responses to redistributive policies. *Journal of Personality and Social Psychology*, *102*(2), 323–336. https://doi.org/10.1037/a0024598

Lucal, B. (1996). Oppression and privilege: Toward a relational conceptualization of race. *Teaching Sociology*, *24*(3), 245. https://doi.org/10.2307/1318739

Lundström, C. (2017). The White side of migration: Reflections on race, citizenship and belonging in Sweden. *Nordic Journal of Migration Research*, *7*(2), 79. https://doi.org/10.1515/njmr-2017-0014

Marmer, E., Marmer, D., Hitomi, L., & Sow, P. (2010). Racism and the image of Africa in German schools and textbooks. *The International Journal of Diversity in Organisations, Communities and Nations*, *10*(5), 1–12. https://doi.org/10.18848/1447-9532/CGP/v10i05/38927

McIntosh, P. (1988). *White privilege and male privilege: A personal account of coming to see correspondence through work in women's studies* (Working Paper No. 189). Wellesley Centers for Women.

Mitchell, G., & Tetlock, P. E. (2009). Disentangling reasons and rationalizations: Exploring perceived fairness in hypothetical societies. In J. T. Jost, A. C. Kay, & H. Thorisdottir (Eds.), *Social and psychological bases of ideology and system justification* (pp. 126–157). Oxford University PressNew York. https://doi.org/10.1093/acprof:oso/9780195320916.003.006

Murdoch, A., & McAloney-Kocaman, K. (2019). Exposure to evidence of White privilege and perceptions of hardships among White UK residents. *Race and Social Problems*, *11*(3), 205–211. https://doi.org/10.1007/s12552-019-09262-3

Nedelsky, N. R. (2023). A New German 'We'? Everyday Perspectives on Germanness and its Boundaries. *Nationalities Papers*, 1–20. https://doi.org/10.1017/nps.2022.106

Nelson, J. C., Adams, G., & Salter, P. S. (2013). The Marley hypothesis: Denial of racism reflects ignorance of history. *Psychological Science*, *24*(2), 213–218. https://doi.org/10.1177/0956797612451466

Nettle, D., & Saxe, R. (2020). Preferences for redistribution are sensitive to perceived luck, social homogeneity, war and scarcity. *Cognition*, *198*, 104234. https://doi.org/10.1016/j.cognition.2020.104234

Nixon, S. A. (2019). The coin model of privilege and critical allyship: Implications for health. *BMC Public Health*, *19*(1), 1637. https://doi.org/10.1186/s12889-019-7884-9

Ogette, T. (2023). *exit RACISM: Rassismuskritisch denken lernen* (11. edition.). Unrast.

Phillips, L. T., & Jun, S. (2021). Why benefiting from discrimination is less recognized as discrimination. *Journal of Personality and Social Psychology*. Advance online publication. https://doi.org/10.1037/pspi0000298

Phillips, L. T., & Lowery, B. S. (2015). The hard-knock life? Whites claim hardships in response to racial inequity. *Journal of Experimental Social Psychology*, *61*, 12–18. https://doi.org/10.1016/j.jesp.2015.06.008

Phillips, L. T., & Lowery, B. S. (2020). I ain't no fortunate one: On the motivated denial of class privilege. *Journal of Personality and Social Psychology*, *119*(6), 1403–1422. https://doi.org/10.1037/pspi0000240

Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology*, *66*, 29–38. https://doi.org/10.1016/j.jesp.2015.09.012

Pratto, F., Çidam, A., Stewart, A. L., Zeineddine, F. B., Aranda, M., Aiello, A., Chryssochoou, X., Cichocka, A., Cohrs, J. C., Durrheim, K., Eicher, V., Foels, R., Górska, P., Lee, I.-C., Licata, L., Liu, J. H., Li, L., Meyer, I., Morselli, D., . . . Henkel, K. E. (2013). Social dominance in context and in individuals. *Social Psychological and Personality Science*, *4*(5), 587–599. https://doi.org/10.1177/1948550612473663

Puryear, C., Steele, L. M., Lawler, J., & Vandello, J. (2019). Reminders of ingroup privilege often threaten identity but rarely change beliefs. https://psyarxiv.com/gvjkc/download?format=pdf

Radke, H. R. M., Kutlaca, M [M.], Siem, B., Wright, S. C., & Becker, J. C. (2020). Beyond allyship: Motivations for advantaged group members to engage in action for disadvantaged groups. *Personality and Social Psychology Review*, *24*(4), 291–315. https://doi.org/10.1177/1088868320918698

Roig, E. (2016). Uttering “race” in colorblind France and post-racial Germany. In K. Fereidooni & M. El (Eds.), *Rassismuskritik und Widerstandsformen* (1st ed., pp. 613–627). Springer VS. https://doi.org/10.1007/978-3-658-14721-1_36

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in Psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, *10*, 813. https://doi.org/10.3389/fpsyg.2019.00813

Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science*, *8*(4), 379–386. https://doi.org/10.1177/1948550617715068

Shuman, E., Johnson, D., Saguy, T., & Halperin, E. (2018). Threat to the group’s image can motivate high identifiers to take action against in-group transgressions. *Personality & Social Psychology Bulletin*, *44*(11), 1523–1544. https://doi.org/10.1177/0146167218768800

Sow, N. (2009). *Deutschland Schwarz Weiß: Der alltägliche Rassismus* (6. Auflage, Taschenbuchausgabe). *Goldmann: Vol. 15575*. Goldmann.

Strupp-Levitsky, M. (2020). *Who’s afraid of Whiteness? White privilege, White racial identity, and the motivated defense of White supremacy* [Doctoral dissertation]. Long Island University, Long Island.

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

Täuber, S., & Moughalian, C. (2022). Collective system-supporting inaction: A conceptual framework of privilege maintenance. *European Journal of Social Psychology*, Article ejsp.2860. Advance online publication. https://doi.org/10.1002/ejsp.2860

Uluğ, Ö. M., & Tropp, L. R. (2021). Witnessing racial discrimination shapes collective action for racial justice: Enhancing awareness of privilege among advantaged groups. *Journal of Applied Social Psychology*, *51*(3), 248–261. https://doi.org/10.1111/jasp.12731

Unzueta, M. M., & Lowery, B. S. (2008). Defining racism safely: The role of self-image maintenance on white Americans’ conceptions of racism. *Journal of Experimental Social Psychology*, *44*(6), 1491–1497. https://doi.org/10.1016/j.jesp.2008.07.011

Weiner, M. F. (2016). Colonized curriculum. *Sociology of Race and Ethnicity*, *2*(4), 450–465. https://doi.org/10.1177/2332649215628037

Zimmerer, J. (2012). Expansion und Herrschaft: Geschichte des europäischen und deutschen Kolonialismus. *Aus Politik Und Zeitgeschichte*, *62*, 10–16. https://www.bpb.de/system/files/dokument_pdf/APuZ_2012-44-45_online_Kolonialismus.pdf

Supplementary Material for

Addressing the "Why" in White Privilege: The Limits of Mere Evidence Confrontations and the Potential of Contextualizing White Privilege in National Colonial History

Section A: Study 1

Results of path analysis with preregistered sample size (N = 220) and without exclusion of failed attention checks

We used the same path analysis procedure with the pre-registered sample size of *N* = 220. The model had acceptable to good fit, $\chi^2(1) = 2.18$, $p = .14$, RMSEA = .10 (95% CI [0.00 0.28], $p = .20$, SRMR = .04, CFI = .99. However, it only explained low amount of variance in the endogenous variables (life hardships: $R^2 = .01$, believe in White privilege: $R^2 = .01$, and acceptance of White privilege: $R^2 = .02$). Parameter estimates are presented in Table 1. The indirect effect of condition through life hardships on acceptance of White privilege was not significant, $b = -.03$ (95% CI [-.14 .04]), $\beta = -.006$, $p = .56$. The results without exclusion of failed attention checks are summarized in Table 2.

Table 1. Parameter estimates for the pre-registered sample size

			95% CI			
Dependent	Predictor	b	Lower	Upper	β	p
Life Hardships	Condition	-0.20	-0.61	0.21	-.08	.348
Believe in White privilege	Condition	0.31	-0.19	0.83	.11	.228
Acceptance of White privilege	Condition	0.45	-0.27	1.18	.11	.231
Acceptance of White privilege	Life Hardships	0.13	-0.10	0.36	.08	.264

Table 2. Parameter estimates for sample without exclusion of failed attention checks

			95% CI			
Dependent	Predictor	b	Lower	Upper	β	p
Life Hardships	Condition	0.04	-0.28	0.36	.02	.824
Believe in White privilege	Condition	0.05	-0.34	0.42	.02	.815
Acceptance of White privilege	Condition	0.00	-0.55	0.53	.00	.999
Acceptance of White privilege	Life Hardships	0.02	-0.15	0.20	.01	.779

Section B: Study 2

Belief in White privilege

Belief in White privilege was measured similarly to the pilot study with only slight changes. We replaced “in Germany” with “in our society” in one item and “unearned privileges” with “privileges (= unearned advantages)” in another (α = .92).

Section C: Study 3

Additional measures

Allyship. We used the Allyship subscale of the Political Solidarity Measure (Neufeld et al., 2019) to tap into feelings of solidarity and allyship with Black people (α = .91). The wording of the first item was changed slightly to: "I feel a sense of attachment to Black people".

Mechanism checks. We used multiple questions to tap into potential mechanisms behind the contextualized confrontation. First, participants were asked to answer to an open question: "What is in your opinion the reason for the different status of White and Black people in our society?". Additionally, we presented a three-item scale to measure White responsibility for racial injustice (Leach et al., 2006). Example item: "I believe that the better position of White people is caused by White people's practices" (α = .82). Finally, we measured emotional appraisal of racial injustice with anger, "It offends me that Black people are not treated as well as White people" (Leach et al., 2006), and guilt, "I feel guilty about the social inequality between White and Black people" (adapted from Swim & Miller, 1999). Results for these variables appear in the extended correlation and means tables below.

Prior engagement with manipulation themes. As additional control questions, we asked participants how much they had engaged with each theme presented in the three conditions prior to their study participation: German voting rights, White privilege and Germany's history of colonialism.

Indirect effects tables

The indirect effects of the preregistered path model appear in Table 3 and of the multiple-group path analysis (split for migration background) are summarized in Table 4.

Table 3. Indirect effects of the preregistered path model

		95% CI			
Description	b	Lower	Upper	β	p
Effects of dummy: contextualized (1) vs. control (0)					
Social Justice Morality ⇒ Solidarity-Based Collective Action Intentions	0.167	-0.045	0.378	.053	.122
Social Justice Morality ⇒ Acceptance of White Privilege	0.151	-0.04	0.342	.039	.122
Life Hardships ⇒ Solidarity-Based Collective Action Intentions	0.008	-0.023	0.04	.003	.597
Life Hardships ⇒ Acceptance of White Privilege	0.064	-0.019	0.147	.016	.132
Effects of dummy: mere evidence (1) vs. control (0)					
Social Justice Morality ⇒ Solidarity-Based Collective Action Intentions	0.062	-0.14	0.263	.019	.55
Social Justice Morality ⇒ Acceptance of White Privilege	0.056	-0.127	0.238	.014	.55
Life Hardships ⇒ Solidarity-Based Collective Action Intentions	0.004	-0.014	0.023	.001	.646

Life Hardships ⇒ Acceptance of White Privilege	0.032	-0.048	0.113	.008	.433

Table 4. Indirect effects of path analysis split for migration background

		95% CI			
Description	b	Lower	Upper	β	p
Effects of dummy: contextualized (1) vs. control (0) for participants with migration background					
Social Justice Morality ⇒ Solidarity-Based Collective Action Intentions	0.423	0.099	0.747	0.135	.01
Social Justice Morality ⇒ Acceptance of White Privilege	0.250	-0.001	0.502	0.068	.051
Life Hardships ⇒ Solidarity-Based Collective Action Intentions	0.000	-0.111	0.112	0.000	.997
Life Hardships ⇒ Acceptance of White Privilege	0.244	0.005	0.483	0.066	.045
Effects of dummy: mere evidence (1) vs. control (0) for participants with migration background					
Social Justice Morality ⇒ Solidarity-Based Collective Action Intentions	0.080	-0.241	0.402	0.024	.624
Social Justice Morality ⇒ Acceptance of White Privilege	0.048	-0.148	0.243	0.012	.632
Life Hardships ⇒ Solidarity-Based Collective Action Intentions	0.000	-0.134	0.134	0.000	.997
Life Hardships ⇒ Acceptance of White Privilege	0.294	0.046	0.542	0.075	.02
Effects of dummy: contextualized (1) vs. control (0) for participants without migration background					
Social Justice Morality ⇒ Solidarity-Based Collective Action Intentions	-0.018	-0.302	0.265	-0.006	.899
Social Justice Morality ⇒ Acceptance of White Privilege	-0.019	-0.31	0.273	-0.005	.899
Life Hardships ⇒ Solidarity-Based Collective Action Intentions	0.004	-0.023	0.031	0.001	.784
Life Hardships ⇒ Acceptance of White Privilege	0.008	-0.048	0.064	0.002	.769
Effects of dummy: mere evidence (1) vs. control (0) for participants without migration background					
Social Justice Morality ⇒ Solidarity-Based Collective Action Intentions	0.044	-0.212	0.299	0.014	.738
Social Justice Morality ⇒ Acceptance of White Privilege	0.045	-0.218	0.308	0.011	.738
Life Hardships ⇒ Solidarity-Based Collective Action Intentions	-0.012	-0.047	0.022	-0.004	.486

Life Hardships ⇒ Acceptance of White Privilege	-0.027	-0.093	0.038	-0.007	.413

Extended table of correlations

Table 5. Table of correlations between all dependent variables

	1	2	3	4	5	6	7	8	9	10	11	12
1 Hardships	—											
2 Social Justice Morality	-.03	—										
3 Privilege Acceptance	-.16**	.39***	—									
4 Collective Action	-.05	.55***	.49***	—								
6 Allyship	-.02	.45***	.29***	.57***	—							
7 White Responsibility	-.13*	.46***	.55***	.49***	.38***	—						
8 Anger at Injustice	-.13*	.49***	.43***	.57***	.53***	.57***	—					
9 Guilt	-.02	.48***	.51***	.57***	.39***	.42***	.39***	—				
10 Engagement Voting	.02	.05	-.05	.09	.14*	-.09	-.07	.01	—			
11 Engagement Privilege	.00	.29***	.43***	.46***	.37***	.19***	.25***	.34***	.26***	—		
12 Engagement Colonialism	.03	.11*	.08	.18***	.15**	-.01	.07	.15**	.43***	.47***	—	
13 Historical Closure	.03	-.44***	-.43***	-.45***	-.25***	-.46***	-.37***	-.40***	.02	-.27***	-.14**	—

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

Extended table of means and standard deviations across conditions

Table 6. Descriptive information and ANOVA results extended

	Condition	N	Mean	SD	ANOVA F (df1,df2)	*p*	η^2_p
Hardships	Control	118	3.77	1.41	1.31 (2,344)	.27	.008
	Contextualized	117	3.48	1.31			
	Evidence	112	3.63	1.40			
Social Justice Morality	Control	118	5.03	1.06	1.20 (2,344)	.30	.007
	Contextualized	117	5.26	1.24			
	Evidence	112	5.11	1.14			
Acceptance of White privilege	Control	117	4.36	1.86	1.65 (2,343)	.19	.010
	Contextualized	117	4.09	1.87			
	Evidence	112	4.52	1.76			
Solidarity-Based Collective Action	Control	118	3.84	1.38	1.84 (2,344)	.16	.011
	Contextualized	117	4.21	1.57			
	Evidence	112	3.98	1.49			
Allyship	Control	118	5.03	1.30	1.61 (2,344)	.20	.009
	Contextualized	117	5.32	1.26			
	Evidence	112	5.07	1.36			
White Responsibility	Control	118	5.40	1.50	1.40 (2,344)	.25	.008
	Contextualized	117	5.70	1.30			
	Evidence	112	5.49	1.29			
Anger at Injustice	Control	118	5.89	1.40	0.28 (2,344)	.76	.002
	Contextualized	117	5.97	1.44			
	Evidence	112	5.84	1.30			
Guilt	Control	118	3.11	1.92	1.14 (2,344)	.32	.007
	Contextualized	117	3.41	2.02			
	Evidence	112	3.06	1.75			
Knowledge Voting	Control	118	4.34	1.55	1.09 (2,344)	.34	.006
	Contextualized	117	4.64	1.56			
	Evidence	112	4.47	1.60			
Knowledge White Privilege	Control	118	4.32	1.76	0.67 (2,344)	.52	.004
	Contextualized	117	4.07	1.72			
	Evidence	112	4.18	1.58			
Knowledge German Colonialism	Control	118	3.94	1.77	0.92 (2,344)	.40	.005
	Contextualized	117	3.68	1.75			

	Evidence	112	3.96	1.62			
Historical Closure	Control Contextualized Evidence	118 117 112	4.19 3.95 4.38	1.34 1.57 1.43	2.54 (2,344)	.08	.02

Comparison between White Germans with and without a migration background

A summary of two-sided Welch's t-tests with migration background (yes / no) as independent and sociodemographic and control variables as outcomes appears in Table 7.

Table 7. T-Tests of migration background on additional variables

	Welch's t	df	p	Cohen's d
Knowledge Voting	-1.68	287.69	.095	-0.18
Knowledge White Privilege	0.92	297.89	.361	0.10
Knowledge German Colonialism	-1.13	281.65	.259	-0.13
Anger at Injustice	-1.05	301.33	.297	-0.11
Guilt	-1.48	307.44	.139	-0.16
Political Orientation	-0.66	294.27	.509	-0.07
Socio-Economic Status	0.69	304.03	.490	0.08
Age	-3.96	344.23	< .001	-0.42
Historical Closure	-0.40	289.60	.686	-0.04

The results for the split path model controlled for age are summarized in Table 8.

Table 8. Summary of path model controlled for age

			95% CI			
Dep	Pred	b	Lower	Upper	β	p
Migration Background						
Social Justice Morality	Dummy1	0.585	0.185	0.986	.266	.004
Social Justice Morality	Dummy2	0.117	-0.323	0.556	.05	.603
Social Justice Morality	Age	0.029	-0.181	0.239	.021	.787
Life Hardships	Dummy1	-0.62	-1.136	-0.103	-.218	.019
Life Hardships	Dummy2	-0.713	-1.263	-0.163	-.237	.011
Life Hardships	Age	0.17	-0.104	0.443	.097	.224
SBCAI	Dummy1	0.337	-0.161	0.835	.108	.185
SBCAI	Dummy2	-0.114	-0.672	0.444	-.035	.689
SBCAI	Social Justice Morality	0.723	0.547	0.899	.509	< .001
SBCAI	Life Hardships	0.000	-0.184	0.184	.000	.999

SBCAI	Age	-0.007	-0.252	0.238	-.004	.954
Privilege Acceptance	Dummy1	-0.261	-0.96	0.438	-.07	.464
Privilege Acceptance	Dummy2	-0.18	-0.867	0.508	-.046	.609
Privilege Acceptance	Social Justice Morality	0.425	0.145	0.705	.253	.003
Privilege Acceptance	Life Hardships	-0.401	-0.613	-0.189	-.308	< .001
Privilege Acceptance	Age	0.097	-0.195	0.389	.043	.515
No Migration Background						
Social Justice Morality	Dummy1	-0.022	-0.435	0.391	-.009	.916
Social Justice Morality	Dummy2	0.061	-0.31	0.432	.024	.747
Social Justice Morality	Age	-0.021	-0.194	0.152	-.019	.811
Life Hardships	Dummy1	-0.09	-0.547	0.366	-.031	.698
Life Hardships	Dummy2	0.239	-0.23	0.708	.083	.317
Life Hardships	Age	0.099	-0.084	0.283	.079	.289
SBCAI	Dummy1	0.105	-0.319	0.529	.033	.627
SBCAI	Dummy2	0.17	-0.208	0.548	.055	.377
SBCAI	Social Justice Morality	0.686	0.558	0.813	.561	< .001
SBCAI	Life Hardships	-0.047	-0.169	0.075	-.044	.449
SBCAI	Age	-0.112	-0.273	0.048	-.083	.17
Privilege Acceptance	Dummy1	-0.603	-1.149	-0.056	-.151	.031
Privilege Acceptance	Dummy2	0.123	-0.437	0.683	.031	.666
Privilege Acceptance	Social Justice Morality	0.705	0.548	0.862	.454	< .001
Privilege Acceptance	Life Hardships	-0.111	-0.272	0.05	-.082	.178
Privilege Acceptance	Age	-0.154	-0.365	0.056	-.09	.15

Note. SBCAI = Solidarity-based collective action intention, Dummy1 = contextualized versus control, Dummy2 = evidence versus control

Section D: Manipulation material used in Study 2 and 3Condition: Contextualized Confrontation

Rassismus wird hierzulande oft mit dem Nationalsozialismus oder Neonazis verbunden. Doch der Rassismus wurde bereits in der Kolonialzeit genutzt, um die Ausbeutung des globalen Südens, insbesondere des afrikanischen Kontinents, zu legitimieren.

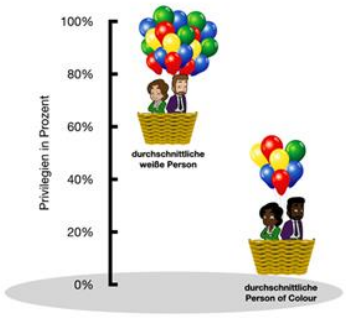
Zwischen 1879 und 1919 bereicherte sich Deutschland materiell und kulturell an seinen Kolonien. Währenddessen verarmte die dortige Bevölkerung, wurde unterdrückt und ermordet. Auch den ersten Genozid des 20. Jahrhunderts verschuldete das Deutschen Reich im damaligen Deutsch-Südwestafrika. Deutsche Truppen töteten dort zwischen 50.000 und 70.000 Herero und Nama.

https://upload.wikimedia.org/wikipedia/commons/b/b4/Herero_and_Nama_prisoners.jpg

Die stereotype Darstellung von schwarzen Menschen als eine minderwertige oder gefährliche Gruppe wurde in der Vergangenheit geschaffen, um den Kolonialismus zu rechtfertigen. Dies hat Auswirkungen bis in die Gegenwart hinein, auch durch die systematische Bevorzugung von weißen Menschen. Diese Bevorzugung äußert sich in der Form von *weißen* Privilegien, z.B. …

- im Alltag, indem *Weiße* einander respektieren und keine Angst voreinander haben oder kriminelle Absichten unterstellen.
- in der Bildung, indem *Weiße* ihre Intelligenz untereinander nicht von vorneherein Infrage stellen.
- auf der Arbeit, indem *Weiße* lieber miteinander arbeiten und sich eher vertrauen.

Durch diese Mechanismen werden die sozialen Unterschiede zwischen *weißen* und nicht-*weißen* Menschen erhalten und verstärkt. Dabei ist wichtig zu beachten, dass dies oft automatisch, also ohne böse Absichten, geschieht. Da Privilegien mit Macht und Einfluss einhergehen, können *weiße* Menschen in Deutschland Verantwortung übernehmen, indem sie versuchen, ungerechte Strukturen zu verändern und ihre Privilegien zu hinterfragen. Auch, um das Erbe des Kolonialismus zu überwinden.

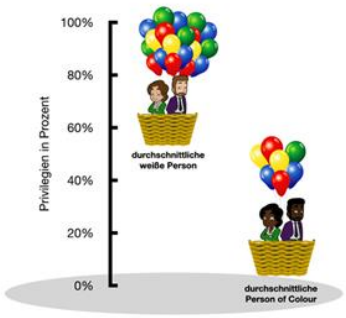


Codition: Evidence Confrontation

Schwarzen Menschen werden oft stereotypisch als eine minderwertige oder gefährliche Gruppe dargestellt. Dies führt zu systematischer Bevorzugung von weißen Menschen. Diese Bevorzugung äußert sich in der Form von *weißen* Privilegien, z.B. …

- im Alltag, indem *Weiße* einander respektieren und keine Angst voreinander haben oder kriminelle Absichten unterstellen.
- in der Bildung, indem *Weiße* ihre Intelligenz untereinander nicht von vorneherein Infrage stellen.
- auf der Arbeit, indem *Weiße* lieber miteinander arbeiten und sich eher vertrauen.

Durch diese Mechanismen werden die sozialen Unterschiede zwischen *weißen* und nicht-*weißen* Menschen erhalten und verstärkt. Dabei ist wichtig zu beachten, dass dies oft automatisch, also ohne böse Absichten, geschieht. Da Privilegien mit Macht und Einfluss einhergehen, können *weiße* Menschen in Deutschland Verantwortung übernehmen, indem sie versuchen, ungerechte Strukturen zu verändern und ihre Privilegien zu hinterfragen.



Condition: Control

In seiner Regierungserklärung vom 28. Oktober 1969 kündigte die SPD/FDP-geführte Bundesregierung an, sie werde dem Bundestag ein Gesetz zur Absenkung des aktiven Wahlalters auf 18 und des passiven Wahlalters auf 21 Jahre vorlegen sowie die Volljährigkeitsgrenze überprüfen.

Bei der Abstimmung am 18. Juni 1970 erreichte die Vorlage die für eine Änderung des Grundgesetzes notwendige Zweidrittelmehrheit. In allen drei Fraktionen – Union, SPD und FDP – gab es am Ende keine Gegenstimmen. Am 31. Juli 1970 trat die Grundgesetzänderung in Kraft.

Heute ist weitgehend unstrittig, dass die Absenkung des Wahlalters auf 18 Jahre die politische Teilhabe junger Menschen erhöht hat. Die Wahlbeteiligung der 18- bis 21-Jährigen war bei den vergangenen Bundestagswahlen allerdings etwas geringer als die durchschnittliche Wahlbeteiligung (um 7,3 Prozent). Bei der Bundestagswahl 2017 stellten die über 60-Jährigen erstmals die größte Wählergruppe dar.

Heutzutage fordern Politikerinnen und Politiker von Grünen, SPD und FDP eine weitere Reduzierung des Wahlalters auf 16 Jahre. Kritikerinnen und Kritiker mahnen hingegen, dass 16- und 17-Jährige leichter manipulierbar seien und aufgrund ihrer Minderjährigkeit noch nicht die nötige Reife hätten. Die Argumente sind auf beiden Seiten vergleichbar mit denen, die der Reform 1970 vorangingen.

References

Leach, C. W., Iyer, A., & Pedersen, A. (2006). Anger and guilt about ingroup advantage explain the willingness for political action. *Personality & Social Psychology Bulletin*, *32*(9), 1232–1245. https://doi.org/10.1177/0146167206289729

Neufeld, K. H. S., Starzyk, K. B., & Gaucher, D. (2019). Political solidarity: A theory and a measure. *Journal of Social and Political Psychology*, *7*(2), 726–765. https://doi.org/10.5964/jspp.v7i2.1058

Swim, J. K., & Miller, D. L. (1999). White guilt: Its antecedents and consequences for attitudes toward affirmative action. *Personality & Social Psychology Bulletin*, *25*(4), 500–515.

Conclusion

Inequality is entrenched, systemic, and resistant to change, because it maintains itself through normalization (Täuber & Moughalian, 2022). To bring about social change, it is therefore important to delegitimize inequality among those who profit from it. The current state of the literature provided only few attempts toward this goal. As we have seen, most research (we also discussed some exceptions) has investigated the outcomes of confrontations with evidence of privilege and focused on self-esteem to explain them. This approach has the potential to individualize privilege as the evidence confrontation is unlikely to raise awareness about the bigger picture and the actual problem: systemic injustice. I argued that to remove the "blinds" on systemic injustice a clearer examination of the role of morality in relation to privilege confrontations is needed. This led to our main question: How can privilege confrontations increase the moral relevance of ingroup advantage?

The manuscripts reported above contribute important evidence toward this question. In both manuscripts we used the 3D model (Knowles et al., 2014) as a starting point, because it provides testable (but so far not thoroughly examined) hypotheses on how people react to confrontations with ingroup advantage. Failing to find support for the main claims of the 3D model, we further focused on the role of morality as an underexplored but potentially decisive factor to understand constructive outcomes of privilege interventions. Manuscript 1 demonstrated that we can separate between moral appraisal and image appraisal of group-based (male) privilege. The evidence showed that moral appraisal of privilege is an important driver behind constructive responses to privilege confrontations and also positive attitudes toward the confronter. In Manuscript 2, we established the important role of moral motives of social justice, not only for the acceptance of White privilege but also for the intention to act in solidarity with Black people. Contextualizing White privilege within Germany's involvement in colonialism could activate these motives, but the effect was limited to White participants who seemed to be more critically

attached to their German identity. Importantly, this is the first time an intervention was reported that affected social justice morality, a construct which has so far been conceptualized as a rather stable characteristic closely related to political ideology (Janoff-Bulman & Carnes, 2016). In total, the manuscripts provide promising leverage points for how privilege interventions can aim beyond the acceptance of individual privilege and toward delegitimizing the status quo of inequality.

I summarize each project's contributions first. Then I will be discussing their results in relation to the 3D model, which brings me to an elaboration of two major empirical problems of the 3D model and to a suggestion for its revision. Afterwards I discuss joint limitations of the empirical basis of the work presented in this thesis. Finally, I will provide a look ahead by summarizing new research questions that flow from this work, which will hopefully inspire scholars to contribute their own perspective and evidence on the important question on how to raise *critical consciousness* among the privileged.

Main empirical take-aways of the research presented in this thesis

In Manuscript 1, we tackled the poorly defined concept of "group-image threat" of the 3D model which did not distinguish clearly between moral and image concerns. As such, we tested a moral appraisal path (consisting of moral threat and moral shame) versus an image appraisal path (consisting of group-image threat and image shame) and assumed that together these would explain respectively positive and negative effects of the privilege interventions. We also introduced backlash against or derogation of the confronter which has not yet been considered as an outcome of privilege interventions. The results confirmed the positive role of moral and the negative role of image appraisal on attitudes toward male privilege and toward the confronter. Furthermore, the appraisal-based model explained more variance in privilege attitudes than did a path analysis with variables based on the 3D model (where meritocratic and group-image threat were used as parallel mediators).

Beyond testing the 3D model, we expected that different intervention styles (confrontation with male privilege versus information about gender inequality) and confronter positionality (male versus female) would affect the relative strength of moral and image appraisal. However, we found that confrontation (versus information) and female (versus male) confronters triggered more image *and* more moral appraisal, which could explain the lack of effects on attitudes overall. Finally, there was strong evidence across the three studies that the female confronter received stronger derogation than the male confronter. We also observed a tendency that men justify male privilege more strongly against a female than against a male confronter.

In Manuscript 2, we conducted three conceptual replications of two similar studies which found that White individuals tend to claim life hardships to defend against confrontations with evidence of their White privilege (Murdoch & McAloney-Kocaman, 2019; L. T. Phillips & Lowery, 2015). The original studies were conducted in the UK and the US, whereas our sample consisted of people with German nationality. Across all three experiments there was no effect of confronting with evidence of White privilege on life hardships, and only in one instance did we find the expected negative association between life hardships and privilege acceptance.

Beyond the 3D model, we examined whether a White privilege intervention can trigger moral motives rooted in the desire to approach social justice. Specifically, we hypothesized that a White privilege confrontation which situates White privilege in Germany's history of colonialism (a contextualized confrontation) would trigger social justice morality and, through that, incite not only privilege acceptance but also political solidarity with Black and People of Color. We found some evidence for the positive effect of a contextualized confrontation, although this effect seemed to be limited to demographics that are open to critically reflect on negative collective memory: White German students, White Germans who do not support historical closure, and White Germans with a migration background.

Empirical problems of the 3D model and a proposed revision

The major claims of the 3D model are that group-image threat motivates the advantaged to dismantle systems of inequality, that meritocratic threat motivates the advantaged to distance from privilege or to deny inequality, and that meritocratic threat is the primary response to ingroup advantage in "individualistic contexts" (Knowles et al., 2014, p. 600). In fact, the authors expect broad generalizability of the whole model not only across different cultural contexts but also across different domains of inequality and position the model as a "[…] a cultural-psychological theory of responses to ingroup dominance […]" (Knowles et al., 2014, p. 604). By holding the model to its own high standards, however, we have uncovered two major empirical problems which cast doubt on all these claims. The first problem concerns the differentiation between group-image threat and moral threat. The second problem concerns the primacy of meritocratic threat and the generalizability of this ideological justification even within Western contexts.

Problem 1: Differentiating between threat to the ingroup's image and to its moral essence

Manuscript 1 provided counterevidence for the alleged positive role of group-image threat: Group-image threat, was consistently associated with *negative* attitudes toward male privilege and toward the confronter. By separating moral from image appraisal, we could also show that concerns for morality but not concerns for the ingroup's image are associated with constructive reactions to ingroup privilege.

As it was, the evidence for a positive effect of group-image threat was already thin. The assumption that experiencing group-image threat would motivate action to dismantle systems of inequality was based on a single study. This study found that among White US individuals, concerns for collective self-esteem increased support for policies which limit the opportunities of White people; however, support for status-enhancing policies for Black people was not affected (Lowery et al., 2012). While we did not investigate specific policy support, we did measure acceptance of feminism as one privilege attitude related to dismantling. Again, moral appraisal

was positively, and image appraisal was negatively related to this outcome. Together Manuscript 1 indicates that moral but not image concerns should be investigated as main motive behind the willingness to dismantle inequality "from above".

Moral and image appraisal also map on the distinction between internalized and symbolic moral identity (Aquino & Reed, 2002). These two dimensions of moral identity serve as differentiation between the *performance* of morality to communicate goodness (symbolic) and an *inherent* moral motivation (internalized), and they have similar relations to intergroup outcomes as moral and image appraisal did in our and in previous research (Allpress et al., 2014). A similar distinction has been made regarding guilt as inward focused and shame as outward focused emotional reaction to identity threat (Leach et al., 2002). The results of Manuscript 1 can therefore be situated well within the literature on moral failure and emotional reactions to injustice which differentiate between a constructive ingroup-focused dimension of threat to morality, connected to feelings of guilt or "moral" shame, and a defensive outgroup-focused dimension of threat to the group's image, connected to feelings of "image" shame or fear of rejection (Cibich et al., 2016; Gausel & Leach, 2011; Gausel et al., 2012). The main problem of the 3D model is that it did not differentiate clearly between these two dimensions of identity threat.

Another important emotional reaction, which the 3D model did not address and which we have not considered so far as well, is outrage at injustice (Leach et al., 2002). Beyond the ingroup/outgroup focus of image and identity, outrage is a morally fueled emotion directed at inequality and is thus associated with action in support for social change (Leach et al., 2006; van de Vyver & Abrams, 2015).

Especially in the realm of morality, it is important to delineate clearly what counts as moral and what does not, because psychological literature tends to infuse "morality" into non-moral concepts to increase the research's suggested relevance (Haslam, 2016). As one recent

investigation of moral emotions demonstrated, when strictly differentiating morality from image concerns, only guilt and outrage remain (Fitouchi et al., 2021). I therefore suggest that the most promising way forward for the field would be to start differentiating more clearly between *moral* emotions based in concerns for the ingroup or inequality (e.g., moral shame, guilt, outrage) and *non-moral* emotions based in concerns for the relation of the ingroup to other groups or to society (e.g., image shame, inferiority, rejection), and how they affect constructive and defensive responses to confrontations with ingroup privilege.

Problem 2: Inequality-justifying ideologies are culture-specific

Both manuscripts failed to confirm the predictions of the 3D model regarding the role of meritocratic and group-image threat: Manuscript 1 found no effect of the confrontation (versus information) on meritocratic threat and Manuscript 2 could not replicate the effect of the confrontation on life hardships. Therefore, in both manuscripts, meritocratic justification of one's status could not be identified as a relevant defensive strategy. It is unclear whether this is due to a different representation of meritocracy (Heuer et al., 2020) or because of a different understanding of racism in Germany (Ball et al., 2022). In any case, the results are in line with the idea that how ideologies shape responses to confrontations with inequality is somewhat culture-specific (Osborne et al., 2017).

The strong focus on US samples in social psychology has been criticized more and more recently. Research from the US and other WEIRD (Western, Educated, Industrialized, Rich, Democratic) contexts often provides the basis for our understanding of the *human* psyche (Henrich et al., 2010). This is obviously concerning in many ways and it is even more worrying as some theories, like the 3D model in this dissertation, can not even be transferred from one WEIRD context to another (i.e., from the USA to Germany). One major reason for this failed replication may lie in the very different colonial history of the US and continental Europe. While the indigenous population was kept largely separated from the European populations, the US

imported masses of Black slaves and has a long and important political history of negotiating the relationship between White and Black Americans. Accordingly, the existence of racism (White privilege and Black oppression) is more visible and more difficult to deny in the US than in Europe (Ball et al., 2022). In contrast to the US, White Germans may simply ignore a confrontation with evidence of their White privilege, because they can be relatively sure of their "privilege of not having to be confrontational while preventing or reversing the consequences of changes brought about by disadvantaged groups for the sake of preserving their power" (Täuber & Moughalian, 2022, p. 7). In other words, there is no need to claim more life hardships or to get upset about a confrontation with evidence of White privilege when not creating much fuzz about it is the most efficient reaction to maintain privilege. Of course, we did not directly compare effects between cultural contexts and much of this discussion therefore borders on speculation and needs to be backed up by more research. But this is exactly the important message: If not specifically investigated, social scientists should rather assume that a phenomenon is shaped by the specific context it was found in than to believe that the human condition materialized to them by chance.

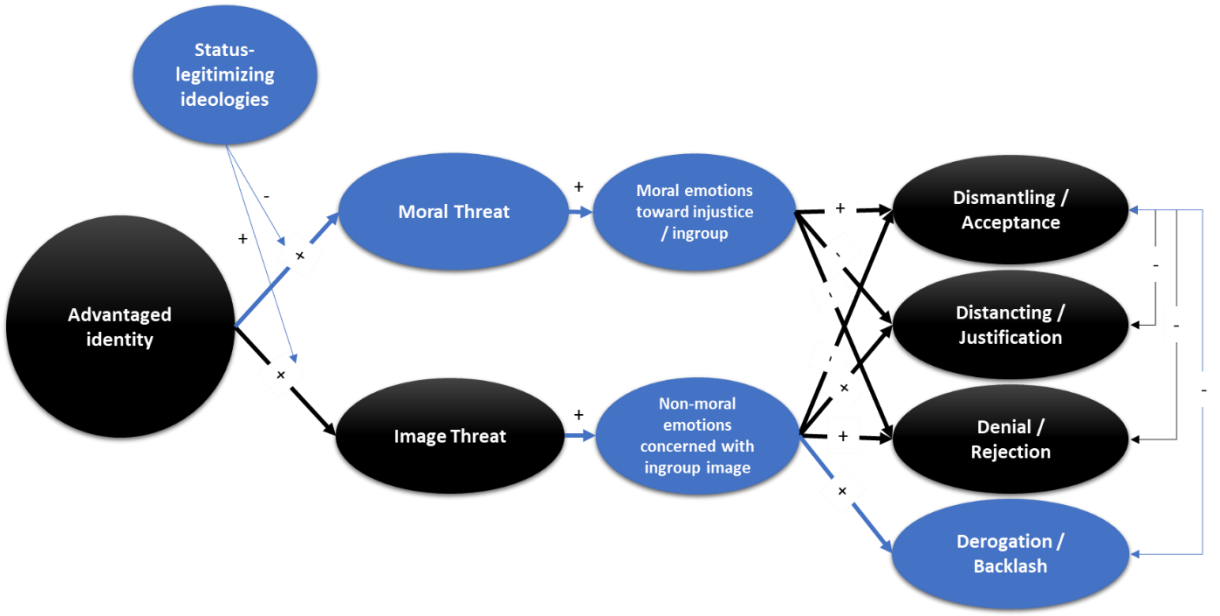
I also need to address one final part of the problem of generalizability: The 3D model claims to be a model about "social dominance" generally and accordingly should be expected to be applicable to systems of inequality beyond racism, such as sexism. However, none of the model's contentions materialized in our research on male privilege. Contrary to the model, I therefore suggest that psychological reactions to ingroup advantage may depend on the available ideological justifications of ingroup advantage in the specific context of inequality (Czopp & Monteith, 2003). For example, meritocratic threat in response to privilege confrontations might be more relevant when the disadvantaged outgroup is stereotyped as lazy, which applies to the stereotype of Black people but not of women, and also not of Black women (Ghavami & Peplau,

2013). This further demonstrates that the assumption of meritocratic threat as a primary psychological response to ingroup advantage is overstated.

A proposed revision of the model

In summary the two manuscripts and the six studies within them point toward the need to adapt and clarify some central predictions of the 3D model: First, group-image threat should be more clearly conceptualized as only about the group's *image* and separated from threat to the *moral* values of the group. Second, image threat should then be expected to have negative effects (leading to more distancing or denial), while moral threat should have positive effects (leading to more dismantling). Third, emotional reactions toward the ingroup and outgroup, and toward inequality might be considered as important motivational forces that connect identity threat with privilege attitudes and support for social change. Fourth, the assumption that meritocratic threat is the primary psychological response to confrontation with ingroup advantage in capitalist cultural contexts is questionable. As such, the availability and nature of status-justifying ideologies might be more generally considered to be a moderator that shapes the appraisal of group privilege. Fifth, derogation of the confronter and their political message is as an important negative outcome of privilege confrontations that has not been considered by the model. We summarize these suggestions in a revised model in Figure 2.

Figure 2. A revised and expanded working model of reactions to group privilege



Note. This proposed revision of the 3D model is based on the findings of the manuscript reported in this thesis. It is not supposed to be a definite new model but a summary of open research questions. Black elements are carried over from the original model because they were supported in our research. Blue elements are new additions that we suggest based on our findings. Moral threat has been split from group-image threat to explain positive reactions to privilege. Meritocratic threat has been replaced by a new proposed moderator describing the availability and nature of status-legitimizing ideologies. It seems likely that for example culturally shared meritocratic beliefs (as well as system justification and other ideologies) shape appraisal of group privilege. Moral and non-moral emotional reactions have been added as potential motivators of the distal outcomes. The new (and old) hypotheses must be tested in additional contexts outside of Germany before the model can truly be called a cultural-psychological model of social dominance.

Joint limitations of the empirical foundation of this thesis

In many ways the two manuscripts improve upon some important methodological limitations of previous research on White and male privilege confrontations. The use of preregistered studies

and sensitivity power analyses as well as the application of mixed methods and sampling within a broad range of demographics were deliberate choices to enhance the internal and external validity of our studies. This is especially valuable regarding the null-findings that challenge some theoretical claims which had been built on marginal results within rather specific samples (Knowles et al., 2014; L. T. Phillips & Lowery, 2015). Still there are some important limitations of the empirical basis of this thesis, specifically regarding the diversity of research methods used, the lack of long-term evaluations of effects, and the missing intersectional analysis.

The lack and value of using more in-depth research methods

Both manuscripts reported in this dissertation relied heavily on quantitative online experiments. The results of these experiments are illuminating in terms of clarifying the role of concepts that have previously not been considered (i.e., morality), may have been emphasized too much (i.e., meritocratic threat), or seem to affect privilege acceptance in the opposite direction than previously expected (i.e., group-image threat). That we were able to find so many inconsistencies in the theoretical approaches that informed the hypotheses and designs of our experiments indicates that our understanding of the psychology of "reflecting on ingroup privilege" is severely limited and urgently needs more in-depth exploration. To broaden our perspective again, we would need more research designs that are able to thoroughly investigate the different understandings of group-based privilege together with their motivational underpinnings without confining people's responses into an arbitrary theoretical corset. Qualitative research designs, such as focus groups or detailed interviews with people from advantaged groups, would for example provide much needed information on the diverse relevant cognitions and emotions in response to group privilege. There are only few such qualitative analyses of attitudes toward group privilege and they provided important information about the motives behind constructive and defensive attitudes toward group-based privilege also to this dissertation (e.g., Ancis & Szymanski, 2001; Eckerle & Lienen, manuscript in preparation).

Another fitting approach would be employing a Q-Method study. Q-Method studies explore the subjective perspectives of a diverse set of respondents instead of determining the validity of the researcher's predetermined "objective" perspective. In them, participants are asked to sort a diverse set of statements on one particular issue (e.g., different opinions on privilege) on a bipolar dimension (e.g., support/reject the opinion) in a rough normal distribution pattern (few extreme and many neutral positions). Through this approach, different "types" (or factors) of people who share a subjective understanding of a certain concept can be extracted. In this way, Q technique "[…] can challenge the power of the researcher over the researched and so allow emancipatory 'surprise' in the process" (Robbins & Krueger, 2000, p. 645). Supporting an emancipatory research agenda with an emancipatory research design has the potential to challenge previous and to uncover new ways to foster political solidarity and social change.

The lack and value of investigating causal mechanisms and long-term effects of privilege interventions

Practical privilege interventions should be able to build upon research that provides some indication about the relevant causal mechanisms and long-term effects of different intervention strategies. In both manuscripts, however, we studied the validity of causal mechanisms based on cross-sectional data, which only allows for an investigation of correlational patterns that *could* be in line with these mechanisms (Fiedler et al., 2011). The manuscripts, as does almost all research on privilege confrontations, also do not provide information on long-term effects and so we can only make cautious claims about the practical use of, for example, embedding White privilege in national colonial history.

Promising evidence comes from an evaluation study of a diversity training which focused on communicating intersectionality and aimed to highlight the complex relationship between privilege and oppression to participants. This training increased privilege acceptance one week later, which causally increased outgroup attitudes another week later (Ehrke et al., 2020). While

the social change potential of increasing outgroup attitudes among the advantaged is limited (Dixon et al., 2010) and would need further investigation, these findings underscore the importance to start moving research from highly controlled experimental designs into the field (Paluck et al., 2021).

The lack and value of addressing intersectionality

Finally, we have not considered the role of intersectionality in our studies. Previous scholars emphasized that studying group privilege confined to one advantaged identity can lead to biased expectations about the actual presence of privilege (Case et al., 2012; Case et al., 2020; Coston & Kimmel, 2012). For example, claims of life hardships might not always be a reaction rooted in meritocratic threat, but they could be an indicator of the existence of group-based disadvantage based on other social identities. In the open answers in Manuscript 1 (Study 2), some men indicated for example that the male privileges listed in the confrontation did not apply to them or that their relation with these privileges was more complicated because of their queer identity (i.e., as homosexual, or because they are not always perceived as male). Thus, male identity should not be equated with participation in hegemonic masculinity as this would misrepresent the diverse manifestations of masculinity in heteronormative societies (Connell, 2020; Pease, 2000). In a similar instance, people who identify as Black or bi-*racial* but are perceived as White (and vice versa) may have a much more complicated relationship between their identity and racism than the research in this thesis was able to investigate.

Taken together, identification with an advantaged social category alone is not enough to determine one's relation with its associated system of inequality. This highlights the limits of (research on) privilege interventions that only focus on one system of inequality. One main take away of this dissertation is that embedding group privilege within complex systems of inequality has the potential to activate moral concerns about injustice which can in turn motivate political solidarity (discussed in more detail in the next section). However, because both manuscripts

disregarded intersectional perspectives on systemic inequality I can only cautiously infer the role of such *critical consciousness* from their findings.

Relatedly, the studies reported here also lack an intersectional perspective in their main outcome variables. Specifically, we did not investigate how privilege confrontations can affect intersectional awareness (awareness of systemic inequality beyond the specific one addressed in the confrontation). Previous research has shown that such intersectional awareness may be an important predictor of critical attitudes toward the status quo and pro-social attitudes toward outgroups (Curtin et al., 2015; Ehrke et al., 2020) as well as for politicization (Curtin et al., 2016). It is important for the social change potential of future research to start adopting intersectional and inclusive approaches more frequently and, hence, to move beyond the intergroup setting and toward examining people's relation to systems of inequality (Boehm, 2022; Chetkovich, 2019; Winker & Degele, 2011).

A look ahead: Confronting the privileged with systemic inequality?

A strong current in this thesis was the idea to broaden our perspective on *how* we confront the advantaged with ingroup privilege. Before I close, I want to summarize some implications of this thesis. This addresses future researchers who are interested in finding out how to effectively moralize inequality, especially among the advantaged.

Past research indicated that moral appraisal of inequality is shielded by social norms rooted in status-defending ideologies, such as social dominance and meritocracy (see also Hartwich & Becker, 2023; Pratto et al., 2013; Täuber & Moughalian, 2022). Efforts to make inequality a moral issue therefore need to pierce through this "ideological armor". We have shown repeatedly that privilege confrontations *can* affect moral considerations: The confrontations (versus information) increased moral appraisal in Manuscript 1 and the contextualized confrontation (versus control) in Manuscript 2 increased social justice morality among samples that seem more

critically attached to their national identity. While these effects were small and full of boundary conditions, they are a promising first step that future research can build on.

One follow-up question may be: How can the impact of privilege confrontations on image threat be reduced while emphasizing the immorality of inequality? Image appraisal seems to be an important psychological motivator for defensive reactions against social change efforts (Teixeira et al., 2020) which is exacerbated by status-legitimizing ideology (Hässler et al., 2019). For example, people with strong meritocratic beliefs may be inclined to see confrontations with their privilege as attacks on the value of their hard work or as an attempt to downplay their contribution to society (e.g., Fortgang, 2014). Manuscript 2 demonstrated that a system-critical perspective on ingroup advantage can provide leverage against such ideological armor, and thus help to moralize inequality as unjust. This indicates that providing contextual information on systemic inequality has the potential to lift ideological defenses and to foster moral engagement (see also Applebaum, 2005; Bonam et al., 2019).

Examples which demonstrate how providing new perspectives and contextual information can inspire social change, circumvent ideological armor, and moralize injustice come from various research strands: For example, heightening intersectional awareness inspired politicization rooted in a system-critical perspective (Curtin et al., 2016), a communication of absolute moral ideals was related to more anger at injustice, efficacy to reduce injustice, solidarity with those affected by the injustice, as well as collective action intentions against global injustice (Barth et al., 2015), and reflecting on negative national history seems to induce a moral growth mindset, which increases support for efforts toward reparation (Lienen, 2023). The diversity of research demonstrating how to inspire action for social change demonstrates that privilege interventions have much to gain by “connecting the dots” between different theoretical strands and fields (Ellemers, 2013). Specifically, such integrated research accounts may come up with new

perspectives on the nature of inequality that have the potential to increase critical consciousness in those who have the privilege to remain uncritical.

In all this, avoiding the tendency of individualizing racism is of great importance. Only if we acknowledge the specific systemic processes behind power relations can we derive any valuable meaning from psychological reactions to them. Critical Whiteness scholars (based in the US) for example have demonstrated how situating and analyzing White privilege as a variable of individual difference contributes to obscuring how White supremacy legitimizes privilege (Bonds & Inwood, 2016). A recent illuminating analysis on racial inequality further highlights the dynamic relationship in which individual racial biases shape and are shaped by systems and institutions. This approach highlights the fact that privilege is only one of many systemic contributors to racial inequality, next to cultural narratives and values, segregated communities, shared stereotypes, and nonverbal messages. These factors operate in "feedback cycles" to maintain racial disparities (Skinner-Dorkenoo et al., 2023). Interventions that aim to reduce inequality and foster social change will need to simultaneously address as many of these leverage points as possible to produce lasting effects.

Concluding remarks

Throughout this thesis I have advocated repeatedly for adopting a meta-theoretical perspective to examine how context affects the (emotional) appraisal of privilege and the moral relevance of social justice. Integrating psychological reactions to ingroup advantage with an analysis of contextual circumstances "accurately reflects the social and psychological dynamics of sentiment, theory and research" (Leach & Teixeira, 2022, p. 26) and thus helps guiding practical interventions (see also Bou Zeineddine & Leach, 2021). Furthermore, such a "system of systems" perspective can help to evaluate research in light of its contribution to social justice (Adams et al., 2008). Take for example the finding cited multiple times above: that privilege

interventions threaten self-esteem, and that this leads to reduced privilege acceptance and less pro-egalitarian action intentions. From a purely psychological perspective, one would take this finding to call for interventions that protect self-esteem (e.g., self-affirmation) thereby removing the negative impact of threatened self-esteem from the equation. From a systems perspective, however, we can see that negative states can be harnessed to *motivate* constructive behavior by shaping their appraisal. For example, shame often leads to withdrawal or defensiveness but if it is based in a moral failure perceived to be repairable, it rather motivates redemption (Leach, 2016). I believe that the impact of context is far from trivial for the research on social justice.

Systems of inequality provide advantages to some groups at the expense of others and injustice thrives when groups are pitted against each other. However, it would be misguided to believe that inequality disappears through mediation (e.g., Dixon et al., 2010). Instead, true liberation potential lies in the moment at which groups on both ends *realize* their collective social power to change the context in their favor. Humans have proven to be very creative in the way they shape and maintain complex societies. This should motivate us to start looking into why we got stuck within a highly unequal way of structuring society – and how to *move* beyond it (Birch & Buskell, 2022; Graeber & Wengrow, 2021).

References

Abrams, D., & Hogg, M. A. (1990). An introduction to the social identity approach. In D. Abrams & M. A. Hogg (Eds.), *Social identity theory: Constructive and critical advances.* Harvester Wheatsheaf.

Adams, G., Biernat, M., Branscombe, N. R., Crandall, C. S., & Wrightsman, L. S. (2008). Beyond prejudice: Toward a sociocultural psychology of racism and oppression. In G. Adams, M. Biernat, N. R. Branscombe, C. S. Crandall, & L. S. Wrightsman (Eds.), *Commemorating Brown: The social psychology of racism and discrimination* (pp. 215–246). American Psychological Association. https://doi.org/10.1037/11681-012

Allen, A. M., & Leach, C. W. (2018). The psychology of Martin Luther King Jr.'s "Creative Maladjustment" at societal injustice and oppression. *Journal of Social Issues*, *74*(2), 317–336. https://doi.org/10.1111/josi.12271

Allpress, J. A., Brown, R., Giner-Sorolla, R., Deonna, J. A., & Teroni, F. (2014). Two faces of group-based shame: Moral shame and image shame differentially predict positive and negative orientations to ingroup wrongdoing. *Personality & Social Psychology Bulletin*, *40*(10), 1270–1284. https://doi.org/10.1177/0146167214540724

Ancis, J. R., & Szymanski, D. M. (2001). Awareness of White privilege among White counseling trainees. *The Counseling Psychologist*, *29*(4), 548–569.

Applebaum, B. (2005). In the name of morality: Moral responsibility, Whiteness and social justice education. *Journal of Moral Education*, *34*(3), 277–290. https://doi.org/10.1080/03057240500206089

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. https://doi.org/10.1037//0022-3514.83.6.1423

Arndt, S. (2020). Weißsein: Die verkannte Strukturkategorie Europas und Deutschlands. In M. M. Eggers, G. Kilomba, P. Piesche, & S. Arndt (Eds.), *Mythen, Masken und Subjekte: Kritische Weißseinsforschung in Deutschland* (4th ed., pp. 24–28). Unrast.

Ball, E., Steffens, M. C., & Niedlich, C. (2022). Racism in Europe: Characteristics and intersections with other social categories. *Frontiers in Psychology*, *13*, 789661. https://doi.org/10.3389/fpsyg.2022.789661

Barajas, S. (2022). Unearned advantages? Redefining privilege in light of childhood. *Children's Geographies*, *20*(1), 24–36. https://doi.org/10.1080/14733285.2021.1912295

Barth, M., Jugert, P., Wutzler, M., & Fritsche, I. (2015). Absolute moral standards and global identity as independent predictors of collective action against global injustice. *European Journal of Social Psychology*, *45*(7), 918–930. https://doi.org/10.1002/ejsp.2160

Bergkamp, J., Martin, A., & Olson, L. (2022). Social privilege: Flipping the coin of inequity. In J. L. Chin (Ed.), *Race and Ethnicity in Psychology Ser. The Psychology of Inequity.* ABC-CLIO LLC.

Bergkamp, J., Olson, L., & Martin, A. (2022). Before allyship: A model of integrating awareness of a privileged social identity. *Frontiers in Psychology*, *13*, 993610. https://doi.org/10.3389/fpsyg.2022.993610

Birch, J., & Buskell, A. (2022). How we got stuck: The origins of hierarchy and inequality. *Mind & Language*, *37*(4), 751–759. https://doi.org/10.1111/mila.12420

Boehm, O. (2022). *Radikaler Universalismus: Jenseits von Identität*. Propyläen.

Bonam, C. M., Nair Das, V., Coleman, B. R., & Salter, P. (2019). Ignoring history, denying racism: Mounting evidence for the Marley hypothesis and epistemologies of ignorance. *Social Psychological and Personality Science*, *10*(2), 257–265. https://doi.org/10.1177/1948550617751583

Bonds, A., & Inwood, J. (2016). Beyond White privilege: Geographies of White supremacy and settler colonialism. *Progress in Human Geography*, *40*(6), 715–733. https://doi.org/10.1177/0309132515613166

Bou Zeineddine, F., & Leach, C. W. (2021). Feeling and thought in collective action on social issues: Toward a systems perspective. *Social and Personality Psychology Compass*, *15*(7). https://doi.org/10.1111/spc3.12622

Brad, L. C. E., Spisz, T. J., & Tanega, C. G. (2019). Does "privilege checking" make us less racist, or more? Generation and political orientation matter. *Race and Social Problems*, *11*(1), 1–14. https://doi.org/10.1007/s12552-018-9246-0

Branscombe, N. R. (1998). Thinking about one's gender group's privileges or disadvantages: Consequences for well-being in women and men. *British Journal of Social Psychology*, *37*, 167–184. https://doi.org/10.1111/j.2044-8309.1998.tb01163.x

Branscombe, N. R., Schmitt, M. T., & Schiffhauer, K. (2007). Racial attitudes in response to thoughts of white privilege. *European Journal of Social Psychology*, *37*(2), 203–215. https://doi.org/10.1002/ejsp.348

Burson, E., & Godfrey, E. B. (2020). Intraminority solidarity: The role of critical consciousness. *European Journal of Social Psychology*, *50*(6), 1362–1377. https://doi.org/10.1002/ejsp.2679

Case, K. A., Iuzzini, J., & Hopkins, M. (2012). Systems of privilege: Intersections, awareness, and applications. *Journal of Social Issues*, *68*(1), 1–10. https://doi.org/10.1111/j.1540-4560.2011.01732.x

Case, K. A., Rios, D., Lucas, A., Braun, K., & Enriquez, C. (2020). Intersectional patterns of prejudice confrontation by White, heterosexual, and cisgender allies. *Journal of Social Issues*, *76*(4), 899–920. https://doi.org/10.1111/josi.12408

Chancel, L., Pikketty, T., Saez, E., & Zucman, G. (2022). *World Inequality Report 2022*. World Inequality Lab. https://wir2022.wid.world

Chen, C. W., & Gorski, P. C. (2015). Burnout in social justice and human rights activists: Symptoms, causes and implications. *Journal of Human Rights Practice*, *7*(3), 366–390. https://doi.org/10.1093/jhuman/huv011

Chetkovich, C. (2019). How non-binary gender definitions confound (already complex) thinking about gender and public policy. *Journal of Public Affairs Education*, *25*(2), 226–252. https://doi.org/10.1080/15236803.2018.1565050

Cibich, M., Woodyatt, L., & Wenzel, M. (2016). Moving beyond "shame is bad": How a functional emotion can become problematic. *Social and Personality Psychology Compass*, *10*(9), 471–483. https://doi.org/10.1111/spc3.12263

Connell, R. (2020). *Masculinities* (Second edition). Routledge, Taylor & Francis Group. https://www.taylorfrancis.com/books/mono/10.4324/9781003116479/masculinities-connell https://doi.org/10.4324/9781003116479

Coston, B. M., & Kimmel, M. (2012). Seeing privilege where it isn't: Marginalized masculinities and the intersectionality of privilege. *Journal of Social Issues*, *68*(1), 97–111.

Curtin, N., Kende, A., & Kende, J. (2016). Navigating multiple identities: The simultaneous influence of advantaged and disadvantaged identities on politicization and activism. *Journal of Social Issues*, *72*(2), 264–285. https://doi.org/10.1111/josi.12166

Curtin, N., Stewart, A. J., & Cole, E. R. (2015). Challenging the status quo: The role of intersectional awareness in activism for social change and pro-social intergroup attitudes. *Psychology of Women Quarterly*, *39*(4), 512–529. https://doi.org/10.1177/0361684315580439

Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality & Social Psychology Bulletin*, *29*(4), 532–544. https://doi.org/10.1177/0146167202250923

Deutscher Bundestag. (January 2023). *Zur Anerkennung kolonialen Unrechts als Völkermord: Intertemporales Völkerrecht im Kontext des deutsch-namibischen Versöhnungsabkommens*. https://www.bundestag.de/resource/blob/935068/06d354ea81fdc64d7dd41c501a785dd6/WD-2-094-22-pdf-data.pdf

Dixon, J. (Ed.). (2012). *Beyond prejudice: Extending the social psychology of conflict, inequality and social change* (1. publ). Cambridge University Press. http://www.loc.gov/catdir/enhancements/fy1208/2012360417-b.html

Dixon, J., Tropp, L. R., Durrheim, K., & Tredoux, C. (2010). "Let Them Eat Harmony". *Current Directions in Psychological Science*, *19*(2), 76–80. https://doi.org/10.1177/0963721410363366

Does, S., Derks, B., & Ellemers, N. (2011). Thou shalt not discriminate: How emphasizing moral ideals rather than obligations increases Whites' support for social equality. *Journal of Experimental Social Psychology*, *47*(3), 562–571. https://doi.org/10.1016/j.jesp.2010.12.024

Does, S., Derks, B., Ellemers, N., & Scheepers, D. (2012). At the heart of egalitarianism. *Social Psychological and Personality Science*, *3*(6), 747–753. https://doi.org/10.1177/1948550612438924

Dörr, J. (2019, March 7). *Ja, alle Männer*. https://www.sueddeutsche.de/kultur/mansplaining-feminismus-maennlichkeit-1.4357864

Drury, B. J., & Kaiser, C. R. (2014). Allies against sexism: The role of men in confronting sexism. *Journal of Social Issues*, *70*(4), 637–652. https://doi.org/10.1111/josi.12083

Eckerle, F., & Lienen, C. S. (manuscript in preparation). *The social representation of privilege among advantaged activists*.

Ehrke, F., Ashoee, A., Steffens, M. C., & Louvet, E. (2020). A brief diversity training: Raising awareness of ingroup privilege to improve attitudes towards disadvantaged outgroups. *International Journal of Psychology*, *55*(5), 732–742. https://doi.org/10.1002/ijop.12665

Ellemers, N. (2013). Connecting the dots: Mobilizing theory to reveal the big picture in social psychology (and why we should do this). *European Journal of Social Psychology*, *43*(1), 1–8. https://doi.org/10.1002/ejsp.1932

Ellemers, N., Pagliaro, S., & Barreto, M. (2013). Morality and behavioural regulation in groups: A social identity approach. *European Review of Social Psychology*, *24*(1), 160–193. https://doi.org/10.1080/10463283.2013.841490

Ellemers, N., & van den Bos, K. (2012). Morality in groups: On the social-regulatory functions of right and wrong. *Social and Personality Psychology Compass*, *6*(12), 878–889. https://doi.org/10.1111/spc3.12001

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, *47*(6), 1231–1236. https://doi.org/10.1016/j.jesp.2011.05.007

Fitouchi, L., André, J.-B., & Baumard, N. (2021). *Are there really so many moral emotions? Carving morality at its functional joints*. https://psyarxiv.com/2phzt/

Flood, M., Dragiewicz, M., & Pease, B. (2021). Resistance and backlash to gender equality. *Australian Journal of Social Issues*, *56*(3), 393–408. https://doi.org/10.1002/ajs4.137

Fortgang, T. (2014, May 2). Why I'll never apologize for my White male privilege. *Time*. https://time.com/85933/why-ill-never-apologize-for-my-white-male-privilege/

Gausel, N., & Leach, C. W. (2011). Concern for self-image and social image in the management of moral failure: Rethinking shame. *European Journal of Social Psychology*, *41*(4), 468–478. https://doi.org/10.1002/ejsp.803

Gausel, N., Leach, C. W., Vignoles, V. L., & Brown, R. (2012). Defend or repair? Explaining responses to in-group moral failure by disentangling feelings of shame, rejection, and inferiority. *Journal of Personality and Social Psychology*, *102*(5), 941–960. https://doi.org/10.1037/a0027233

Ghavami, N., & Peplau, L. A. (2013). An intersectional analysis of gender and ethnic stereotypes. *Psychology of Women Quarterly*, *37*(1), 113–127. https://doi.org/10.1177/0361684312464203

Goodman, D. J. (2011). *Promoting diversity and social justice: Educating people from privileged groups* (Second edition). *Teaching/Learning Social Justice Ser*. Routledge. http://gbv.eblib.com/patron/FullRecord.aspx?p=683962

Goodman, D. J. (2015). Oppression and privilege: Two sides of the same coin. *Journal of Intercultural Communication*, *18*, 1–14.

Gorski, P. C. (2019). Racial battle fatigue and activist burnout in racial justice activists of color at predominately White colleges and universities. *Race Ethnicity and Education*, *22*(1), 1–20. https://doi.org/10.1080/13613324.2018.1497966

Graeber, D., & Wengrow, D. (2021). *The dawn of everything: A new history of humanity* (First American edition). *FSG: Vol. 75*. Farrar Straus and Giroux.

Gulker, J. E., Mark, A. Y., & Monteith, M. J. (2013). Confronting prejudice: The who, what, and why of confrontation effectiveness. *Social Influence*, *8*(4), 280–293. https://doi.org/10.1080/15534510.2012.736879

Hartwich, L., & Becker, J. C. (2023). The role of ideological and intergroup mechanisms in predicting opposition to redistribution and discrimination against the lower social class. *Journal of Social and Political Psychology*, *11*(1), 167–181. https://doi.org/10.5964/jspp.7171

Haslam, N. (2016). Concept creep: Psychology's expanding notions of harm and their moral basis. In J. P. Forgas, L. J. Jussim, & P. A. M. van Lange (Eds.), *Sydney Symposium of Social Psychology series: Vol. 18. The social psychology of morality*. Routledge.

Hässler, T., Shnabel, N., Ullrich, J., Arditti-Vogel, A., & SimanTov-Nachlieli, I. (2019). Individual differences in system justification predict power and morality-related needs in advantaged and disadvantaged groups in response to group disparity. *Group Processes & Intergroup Relations*, *22*(5), 746–766. https://doi.org/10.1177/1368430218773403

Haug, F. (2008). *Die Vier-in-einem-Perspektive: Politik von Frauen für eine neue Linke* (4. Auflage, Deutsche Originalausgabe). Argument.

Hearn, M. C. (2012). Positionality, intersectionality, and power: Socially locating the higher education teacher in multicultural education. *Multicultural Education Review*, *4*(2), 38–59. https://doi.org/10.1080/2005615X.2011.11102893

Helms, J. E. (2017). The challenge of making Whiteness visible: Reactions to four Whiteness articles. *The Counseling Psychologist*, *45*(5), 717–726. https://doi.org/10.1177/0011000017718943

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2-3), 61-83; discussion 83-135. https://doi.org/10.1017/S0140525X0999152X

Heuer, J.-O., Lux, T., Mau, S., & Zimmermann, K. (2020). Legitimizing inequality. *Comparative Sociology*, *19*(4-5), 542–584. https://doi.org/10.1163/15691330-BJA10017

Iyer, A., Leach, C. W., & Crosby, F. J. (2003). White guilt and racial compensation: The benefits and limits of self-focus. *Personality & Social Psychology Bulletin*, *29*(1), 117–129. https://doi.org/10.1177/0146167202238377

Janoff-Bulman, R., & Carnes, N. C. (2016). Social justice and social order: Binding moralities across the political spectrum. *PloS One*, *11*(3), e0152479. https://doi.org/10.1371/journal.pone.0152479

Johnson, A. G. (2006). *Privilege, power, and difference* (2nd ed.). McGraw-Hill. http://www.loc.gov/catdir/enhancements/fy0619/2004064987-d.html

Jost, J., & Hunyady, O. (2003). 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

Kahalon, R., Shnabel, N., Halabi, S., & SimanTov-Nachlieli, I. (2019). Power matters: The role of power and morality needs in competitive victimhood among advantaged and disadvantaged groups. *British Journal of Social Psychology*, *58*, 452-472. https://doi.org/10.1111/bjso.12276

Kelliher, D. (2018). Historicising geographies of solidarity. *Geography Compass*, *12*(9), e12399. https://doi.org/10.1111/gec3.12399

King, J. E. (1991). Dysconscious racism: Ideology, identity, and the miseducation of teachers. *The Journal of Negro Education*, *60*(2), 133. https://doi.org/10.2307/2295605

Knowles, E. D., Lowery, B. S., Chow, R. M., & Unzueta, M. M. (2014). Deny, distance, or dismantle? How White Americans manage a privileged identity. *Perspectives on Psychological Science*, *9*(6), 594–609. https://doi.org/10.1177/1745691614554658

Kutlaca, M., Becker, J., & Radke, H. R. M. (2020). A hero for the outgroup, a black sheep for the ingroup: Societal perceptions of those who confront discrimination. *Journal of Experimental Social Psychology*, *88*, 103832. https://doi.org/10.1016/j.jesp.2019.103832

Kutlaca, M., Radke, H. R. M., Iyer, A., & Becker, J. C. (2020). Understanding allies' participation in social change: A multiple perspectives approach. *European Journal of Social Psychology*, *50*(6), 1248–1258. https://doi.org/10.1002/ejsp.2720

Lane, A. M., Beedie, C. J., Devonport, T. J., & Friesen, A. P. (2021). Considerations of control groups: Comparing active-control with no treatment for examining the effects of brief intervention. *Sports*, *9*(11). https://doi.org/10.3390/sports9110156

Leach, C. W. (2016). The meta-theory of examining emotion in social relationships. *Psychological Inquiry*, *27*(2), 113–116. https://doi.org/10.1080/1047840X.2016.1162129

Leach, C. W. (2019). After motivational hedonism: Feeling bad can be good | feeling good can be bad. In D. Bain, M. Brady, & J. Corns (Eds.), *Philosophy of suffering: Metaphysics, value, and normativity* (pp. 180–193). Routledge Taylor & Francis Group.

Leach, C. W., & Bou Zeineddine, F. (2021). A systems view of emotion in socio-political context. *Affective Science*, *2*(4), 353–362. https://doi.org/10.1007/s42761-021-00051-z

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. Competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, *93*(2), 234–249. https://doi.org/10.1037/0022-3514.93.2.234

Leach, C. W., Iyer, A., & Pedersen, A. (2006). Anger and guilt about ingroup advantage explain the willingness for political action. *Personality & Social Psychology Bulletin*, *32*(9), 1232–1245. https://doi.org/10.1177/0146167206289729

Leach, C. W., Snider, N., & Iyer, A. (2002). "Poisoning the conscious of the fortunate": The experience of relative advantage and support for social equality. In I. Walker & H. J. Smith (Eds.), *Relative Deprivation* (pp. 136–163). Cambridge University Press.

Leach, C. W., & Teixeira, C. P. (2022). Understanding sentiment toward "Black Lives Matter". *Social Issues and Policy Review*, *16*(1), 3–32. https://doi.org/10.1111/sipr.12084

Lienen, C. S. (2023). *Remembering past ingroup transgressions: Social psychological implications of including negative history in the identity narrative* [Doctoral dissertation]. Philipps-Universität Marburg, Marburg.

Lorde, A. (2012). *Sister Outsider: Essays and Speeches*. Potter/Ten Speed/Harmony/Rodale. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5337027

Lowery, B. S., Chow, R. M., Knowles, E. D., & Unzueta, M. M. (2012). Paying for positive group esteem: How inequity frames affect whites' responses to redistributive policies. *Journal of Personality and Social Psychology*, *102*(2), 323–336. https://doi.org/10.1037/a0024598

Lowery, B. S., Knowles, E. D., & Unzueta, M. M. (2007). Framing inequity safely: Whites' motivated perceptions of racial privilege. *Personality & Social Psychology Bulletin*, *33*(9), 1237–1250. https://doi.org/10.1177/0146167207303016

Lucal, B. (1996). Oppression and privilege: Toward a relational conceptualization of race. *Teaching Sociology*, *24*(3), 245. https://doi.org/10.2307/1318739

Marques, J. M., & Paez, D. (1994). The 'Black Sheep Effect': Social categorization, rejection of ingroup deviates, and perception of group variability. *European Review of Social Psychology*, *5*(1), 37–68. https://doi.org/10.1080/14792779543000011

McIntosh, P. (2012). Reflections and future directions for privilege studies. *Journal of Social Issues*, *68*(1), 194–206. https://doi.org/10.1111/j.1540-4560.2011.01744.x

Minson, J. A., & Monin, B. (2012). Do-gooder derogation: Disparaging morally motivated minorities to defuse anticipated reproach. *Social Psychological and Personality Science*, *3*(2), 200–207. https://doi.org/10.1177/1948550611415695

Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology*, *95*(1), 76–93. https://doi.org/10.1037/0022-3514.95.1.76

Moscovici, S., & Mugny, G. (1983). Minority influence. In P. B. Paulus (Ed.), *Springer Series in Social Psychology. Basic Group Processes* (pp. 41–64). Springer. https://doi.org/10.1007/978-1-4612-5578-9_3

Murdoch, A., & McAloney-Kocaman, K. (2019). Exposure to evidence of White privilege and perceptions of hardships among White UK residents. *Race and Social Problems*, *11*(3), 205–211. https://doi.org/10.1007/s12552-019-09262-3

Nettle, D., & Saxe, R. (2020). Preferences for redistribution are sensitive to perceived luck, social homogeneity, war and scarcity. *Cognition*, *198*, 104234. https://doi.org/10.1016/j.cognition.2020.104234

Nixon, S. A. (2019). The coin model of privilege and critical allyship: Implications for health. *BMC Public Health*, *19*(1), 1637. https://doi.org/10.1186/s12889-019-7884-9

Noelle-Neumann, E. (1974). The spiral of silence: A theory of public opinion. *Journal of Communication*, *24*(2), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

OECD. (2017). *The pursuit of gender equality*. OECD Publishing. https://doi.org/10.1787/9789264281318-en

Osborne, D., Yogeeswaran, K., & Sibley, C. G. (2017). Culture-specific ideologies undermine collective action support: Examining the legitimizing effects of postcolonial belief systems. *Group Processes & Intergroup Relations*, *20*(3), 333–349. https://doi.org/10.1177/1368430216682352

Ozden, J. (2022, April 14). Protest movements could be more effective than the best charities. *Stanford Social Innovation Review*. https://ssir.org/articles/entry/protest_movements_could_be_more_effective_than_the_best_charities

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual Review of Psychology*, *72*, 533–560. https://doi.org/10.1146/annurev-psych-071620-030619

Pease, B. (2000). *Recreating men: Postmodern masculinity politics*. SAGE. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=518775

Pease, B. (2021). The rise of angry white men: Resisting populist masculinity and the backlash against gender equality. In C. Noble & G. F. Ottmann (Eds.), *The challenge of right-wing nationalist populism for social work: A human rights approach* (pp. 55–68). Routledge.

Phillips, L. T., & Lowery, B. S. (2015). The hard-knock life? Whites claim hardships in response to racial inequity. *Journal of Experimental Social Psychology*, *61*, 12–18. https://doi.org/10.1016/j.jesp.2015.06.008

Phillips, L. T., & Lowery, B. S. (2017). Herd invisibility: The psychology of racial privilege. *Current Directions in Psychological Science*, *27*(3), 156–162. https://doi.org/10.1177/0963721417753600

Phillips, L. T., & Lowery, B. S. (2020). I ain't no fortunate one: On the motivated denial of class privilege. *Journal of Personality and Social Psychology*, *119*(6), 1403–1422. https://doi.org/10.1037/pspi0000240

Phillips, N. L., Adams, G., & Salter, P. S. (2015). Beyond adaptation: Decolonizing approaches to coping with oppression. *Journal of Social and Political Psychology*, *3*(1), 365–387. https://doi.org/10.5964/jspp.v3i1.310

Piketty, T. (2022). *Eine kurze Geschichte der Gleichheit* (S. Lorenzer, Trans.). C.H. Beck. https://www.beck-elibrary.de/10.17104/9783406791000 https://doi.org/10.17104/9783406791000

Powell, A. A., Branscombe, N. R., & Schmitt, M. T. (2005). Inequality as ingroup privilege or outgroup disadvantage: The impact of group focus on collective guilt and interracial attitudes. *Personality & Social Psychology Bulletin*, *31*(4), 508–521. https://doi.org/10.1177/0146167204271713

Pratto, F., Çidam, A., Stewart, A. L., Zeineddine, F. B., Aranda, M., Aiello, A., Chryssochoou, X., Cichocka, A., Cohrs, J. C., Durrheim, K., Eicher, V., Foels, R., Górska, P., Lee, I.-C., Licata, L., Liu, J. H., Li, L., Meyer, I., Morselli, D., . . . Henkel, K. E. (2013). Social dominance in context and in individuals. *Social Psychological and Personality Science*, *4*(5), 587–599. https://doi.org/10.1177/1948550612473663

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., & Stewart, A. L. (2012). Group dominance and the half-blindness of privilege. *Journal of Social Issues*, *68*(1), 28–45. https://doi.org/10.1111/j.1540-4560.2011.01734.x

Puryear, C., Steele, L. M., Lawler, J., & Vandello, J. (2019). Reminders of ingroup privilege often threaten identity but rarely change beliefs. https://psyarxiv.com/gvjkc/download?format=pdf

Radke, H. R. M., Kutlaca, M., Siem, B., Wright, S. C., & Becker, J. C. (2020). Beyond allyship: Motivations for advantaged group members to engage in action for disadvantaged groups. *Personality and Social Psychology Review*, *24*(4), 291–315. https://doi.org/10.1177/1088868320918698

Ragaru, N. (2017a). Nationalization through internationalization. Writing, remembering, and commemorating the holocaust in Macedonia and Bulgaria after 1989. *Südosteuropa*, *65*(2). https://doi.org/10.1515/soeu-2017-0019

Ragaru, N. (2017b, March 15). *Contrasting destinies: The plight of Bulgarian Jews and the Jews in Bulgarian-occupied Greek and Yugoslav territories during World War Two.* SciencePo. http://bo-k2s.sciences-po.fr/mass-violence-war-massacre-resistance/en/document/contrasting-destinies-plight-bulgarian-jews-and-jews-bulgarian-occupied-greek-and-yugoslav-

Reicher, S., Cassidy, C., Wolpert, I., Hopkins, N., & Levine, M. (2006). Saving Bulgaria's Jews: an analysis of social identity and the mobilisation of social solidarity. *European Journal of Social Psychology*, *36*(1), 49–72. https://doi.org/10.1002/ejsp.291

Robbins, P., & Krueger, R. (2000). Beyond bias? The promise and limits of Q Method in human geography. *The Professional Geographer*, *52*(4), 636–648. https://doi.org/10.1111/0033-0124.00252

Schmitt, M., Behner, R., Montada, L., Müller, L., & Müller-Fohrbrodt, G. (2000). Gender, ethnicity, and education as privileges: Exploring the generalizability of the existential guilt reaction. *Social Justice Research*, *13*(4), 313–337. https://doi.org/10.1023/A:1007640925819

Sherman, D. K., & Cohen, G. L. (2002). Accepting threatening information: Self-affirmation and the reduction of defensive biases. *Current Directions in Psychological Science*, *11*(4), 119–123.

Shuman, E., Saguy, T., van Zomeren, M., & Halperin, E. (2021). Disrupting the system constructively: Testing the effectiveness of nonnormative nonviolent collective action. *Journal of Personality and Social Psychology*, *121*(4), 819–841. https://doi.org/10.1037/pspi0000333

Sidanius, J., & Veniegas, R. C. (2000). Gender and race discrimination: The interactive nature of disadvantage. In S. Oskamp (Ed.), *Claremont Symposium on Applied Social Psychology Series. Reducing Prejudice and Discrimination* (pp. 57–80). Taylor and Francis.

Siem, B., Oettingen, M. von, Mummendey, A., & Nadler, A. (2013). When status differences are illegitimate, groups' needs diverge: Testing the needs-based model of reconciliation in contexts of status inequality. *European Journal of Social Psychology*, *43*(2), 137–148. https://doi.org/10.1002/ejsp.1929

Skinner-Dorkenoo, A. L., George, M., Wages, J. E., Sánchez, S., & Perry, S. P. (2023). A systemic approach to the psychology of racial bias within individuals and society. *Nature Reviews Psychology.* Advance online publication. https://doi.org/10.1038/s44159-023-00190-z

Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass*, *4*(4), 267–281. https://doi.org/10.1111/j.1751-9004.2010.00254.x

Smedley, A., & Smedley, B. D. (2005). Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race. *The American Psychologist*, *60*(1), 16–26. https://doi.org/10.1037/0003-066X.60.1.16

Spanierman, L. B., & Smith, L. (2017). Roles and responsibilities of White allies: Implications for research, teaching, and practice. *The Counseling Psychologist*, *45*(5), 606–617. https://doi.org/10.1177/0011000017717712

Steele, C. M. (1988). The Psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 261–302). Academic Press. https://doi.org/10.1016/S0065-2601(08)60229-4

Strupp-Levitsky, M. (2020). *Who's afraid of Whiteness? White privilege, White racial identity, and the motivated defense of White supremacy* [Doctoral dissertation]. Long Island University, Long Island.

Subašić, E., Reynolds, K. J., & Turner, J. C. (2008). The political solidarity model of social change: Dynamics of self-categorization in intergroup power relations. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *12*(4), 330–352. https://doi.org/10.1177/1088868308323223

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199–214. https://doi.org/10.1037/0022-3514.68.2.199

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

Täuber, S., & Moughalian, C. (2022). Collective system-supporting inaction: A conceptual framework of privilege maintenance. *European Journal of Social Psychology*, Article ejsp.2860. Advance online publication. https://doi.org/10.1002/ejsp.2860

Teixeira, C. P., Spears, R., & Yzerbyt, V. Y. (2020). Is Martin Luther King or Malcolm X the more acceptable face of protest? High-status groups' reactions to low-status groups' collective action. *Journal of Personality and Social Psychology*, *118*(5), 919–944. https://doi.org/10.1037/pspi0000195

Teo, T. (2015). Critical psychology: A geography of intellectual engagement and resistance. *The American Psychologist*, *70*(3), 243–254. https://doi.org/10.1037/a0038727

Thomas, E. F., McGarty, C., Louis, W. R., Wenzel, M., Bury, S., & Woodyatt, L. (2022). It's about time! Identifying and explaining unique trajectories of solidarity-based collective action to support people in developing countries. *Personality & Social Psychology Bulletin*, *48*(10), 1451–1464. https://doi.org/10.1177/01461672211047083

Thomas, E. F., McGarty, C., & Mavor, K. I. (2009). Transforming "apathy into movement": The role of prosocial emotions in motivating action for social change. *Personality and Social Psychology Review*, *13*(4), 310–333. https://doi.org/10.1177/1088868309343290

Thrift, E., & Sugarman, J. (2019). What is social justice? Implications for psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(1), 1–17. https://doi.org/10.1037/teo0000097

Tittler, M. V., Luoma, J. B., Wade, N. G., & Wei, M. (2023). Effects of a White privilege framing and a test of moderators. *Current Psychology*, 1–13. https://doi.org/10.1007/s12144-023-04671-5

Uluğ, Ö. M., & Tropp, L. R. (2021). Witnessing racial discrimination shapes collective action for racial justice: Enhancing awareness of privilege among advantaged groups. *Journal of Applied Social Psychology*, *51*(3), 248–261. https://doi.org/10.1111/jasp.12731

Unzueta, M. M., & Lowery, B. S. (2008). Defining racism safely: The role of self-image maintenance on white Americans' conceptions of racism. *Journal of Experimental Social Psychology*, *44*(6), 1491–1497. https://doi.org/10.1016/j.jesp.2008.07.011

van de Vyver, J., & Abrams, D. (2015). Testing the prosocial effectiveness of the prototypical moral emotions: Elevation increases benevolent behaviors and outrage increases justice behaviors. *Journal of Experimental Social Psychology*, *58*, 23–33. https://doi.org/10.1016/j.jesp.2014.12.005

van der Toorn, J., Ellemers, N., & Doosje, B. (2015). The threat of moral transgression: The impact of group membership and moral opportunity. *European Journal of Social Psychology*, *45*(5), 609–622. https://doi.org/10.1002/ejsp.2119

van Zomeren, M., Kutlaca, M., & Turner-Zwinkels, F. (2018). Integrating who "we" are with what "we" (will not) stand for: A further extension of the Social Identity Model of Collective Action. *European Review of Social Psychology*, *29*(1), 122–160. https://doi.org/10.1080/10463283.2018.1479347

Vitriol, J. A., & Moskowitz, G. B. (2021). Reducing defensive responding to implicit bias feedback: On the role of perceived moral threat and efficacy to change. *Journal of Experimental Social Psychology*, *96*, 104165. https://doi.org/10.1016/j.jesp.2021.104165

Waterstradt, D. (2018). *Prozess-Soziologie der Elternschaft* [Dissertation]. Pädagogische Hochschule Karlsruhe. http://phka.bsz-bw.de/frontdoor/index/index/docId/118

Wenzel, M., Woodyatt, L., & McLean, B. (2020). The effects of moral/social identity threats and affirmations on psychological defensiveness following wrongdoing. *The British Journal of Social Psychology*, *59*(4), 1062–1081. https://doi.org/10.1111/bjso.12378

Winker, G., & Degele, N. (2011). Intersectionality as multi-level analysis: Dealing with social inequality. *European Journal of Women's Studies*, *18*(1), 51–66. https://doi.org/10.1177/1350506810386084

Appendix

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig, ohne unerlaubte Hilfe Dritter angefertigt und keine anderen als die in der Dissertation angegebenen Hilfsmittel benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Dritte waren an der inhaltlich-materiellen Erstellung der Dissertation nicht beteiligt; insbesondere habe ich hierfür nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren des Autors verwendet worden. Mit dem Einsatz von Software zur Erkennung von Plagiaten bin ich einverstanden.

Marburg, 31.05.2023

Frank Eckerle

Percentage distribution of authorship

Manuscript 1: Appraisal of male privilege

Eckerle, F.: 70%

Rothers, A.: 15% (co-developing idea)

Kutlaca, M.: 6% (close feedback to the manuscript and Study 3)

Henss, L.: 3% (Study 1 was Master thesis)

Agunyego, W.: 3% (Study 1 was Bachelor thesis)

Cohrs, J. C.: 3% (conceptual feedback)

Manuscript 2: Addressing the "Why" in White privilege

Eckerle, F.: 80%

Lienen, C. S.: 10% (co-developing idea, minor writing on collective memory)

Kutlaca, M.: 4% (close feedback to the manuscript)

Kotzur, P. F.: 3% (conceptual feedback)

Cohrs, J. C.: 3% (conceptual feedback)