Flinders University

Understanding the factors that influence men's support and resistance to the movement for gender equality

Anna Barron
BPsyc (Hons)
Flinders University

A thesis submitted to Flinders University, South Australia, in fulfilment of the requirements for the Doctor of Philosophy (Psychology)

College of Education, Psychology and Social Work

November 2022

Table of Contents

Summary

In recent years, movements such as #MeToo and Black Lives Matter have brought global attention to longstanding issues of inequality between groups. Notably, these movements have called on members of the advantaged group to acknowledge their group's power, privilege, and history of perpetrating harm against members of the disadvantaged group. While some advantaged group members have joined the fight to address inequality, others have responded with backlash and resistance (e.g., men who claim that the #MeToo movement discriminates against their group).

The current thesis is guided by two key questions: (1) What underlies advantaged group members' support for (versus their opposition to) movements advocating for progressive social change? and (2) How can we reduce backlash and mobilise advantaged group members in efforts to challenge injustice and inequality? I examine these questions in the context of men's reactions to the movement for gender equality. Specifically, I take a social identity approach to argue that the nature of men's opposition to the movement for gender equality may be shaped by their unique identity-based needs as members of a structurally advantaged group that has been accused of wrongdoing. I further argue that opinion-based identities grounded in support for addressing gender inequality may be particularly well placed to overcome men's resistance to feminist efforts and, in turn, foster their commitment to act for gender equality.

Across eight studies using a combination of experimental paradigms, big data analytics (topic modelling Twitter data), and natural language processing, I explore how men encounter, support, or resist the movement to address gender inequality. Firstly, I provide an analysis of the nature of online support and resistance to the movement for gender equality in the context of the #MeToo movement within Australia (Chapter 2). Secondly, I examine the influence of men's social identification and women's accusations of victimisation on men's

need for morality, defensiveness, and collective action intentions (for both women and men's rights; Chapter 3). Lastly, I explore a method designed to mitigate defensiveness and engage men in efforts to act for gender equality (Chapter 4).

Overall, the findings suggest that men's defensiveness regarding the issue of gender-based violence may not only undermine their intentions to act for women's rights, but that it may also facilitate their intentions to advocate for the rights of their own (privileged) group. However, engaging men in group discussion regarding strategies to address gender inequality may provide a means of overcoming defensive reactions and boosting their commitment to act to address violence against women. This thesis provides important implications for both theory and practice regarding the factors that may promote (or hinder) advantaged group members' commitment to act for equality. Namely, it emphasises the importance of examining the nature of both intra- *and* intergroup discussion (that is, discussions both within *and* between groups) to understand how advantaged group members' reactions to social change are shaped.

Declaration

I certify that this thesis:

1. Does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and

2. The research within will not be submitted for any other future degree or diploma without the permission of Flinders University; and

3. To the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Additionally, I confirm that I received an Australian Government Research Training Program Scholarship to support the completion of this thesis (2019-2022).

Anna Barron

August 2022

Acknowledgements

I am privileged to have had a whole host of people who have supported me to complete this PhD. You are all, objectively, the best, and I am endlessly thankful for your love and support – particularly during these last few months in which I have been frantic with deadline anxiety.

Thank you to Lydia Woodyatt, who since 2016 has opened so many doors for me and has been there to (sometimes quite literally) guide me through them. I could never have imagined starting a PhD (and everything that has transpired since) if it were not for your encouragement and support, and for always seeing things in me that I am not yet able to see.

Thank you to Emma Thomas, who is not only a brilliant genius but also the most kind and caring. Your unflagging support, generosity, advice, patience, and enthusiasm have been instrumental in this achievement. Thank you for seeing what this thesis could be and for helping me to get it there.

My sincerest thank you to Michael Wenzel, for your always thoughtful and generous feedback, care, and support.

Thank you to Paul Williamson, for always cheering me on and for buying everyone deep-fried cauliflower at the Tav on Fridays.

Thank you to past and present members of the PsycJEM lab, the SISC lab, and the broader Flinders psychology crew, who have created such a supportive environment to work and research in. I am especially grateful to Mel Takarangi for always providing such genuine care and support to PhD students and ECRs. Thank you also to the brilliant Mel de Vel-Palumbo for always telling it as it is and making me laugh for doing so. Special gratitude to Ella Moeck and Mikaela Cibich – who I wish I could be like when I grow up.

To Bellie, Christiana, Tess, and Bethany, who have made coming onto campus every day an utter hoot and a joy. You are all powerhouses and I am so thankful for your love and support.

To Clauds, who always makes me feel smarter and funnier by virtue of being The Smartest and The Funniest.

To Myf: I'm so happy that you moved back to Adelaide in February 2020 just before everything went tits up. Thank you for calling me in the final hour and providing me with your characteristic wisdom and humour.

To Katie, thank you for always listening and always knowing exactly what to say.

To Nick and Hannah, thank you for making the last three and a bit years the happiest of my young life. Thank you for the coffees, the wines, the dancing in the kitchen, and all of the sitting on the couch we have done together. Thank you for always making me laugh and putting up with me when I am being ridiculous. You are the most brilliant and silly and I am so lucky to know you. I'm so excited to live with you and Zenni dog until we are all very old and grey.

Last and best, to my most favourite people in the world to whom I owe everything: Mum, Dad, and Cat. I love you very much.

Statement of Co-Authorship

This thesis describes original research undertaken in the College of Education, Psychology, and Social Work at Flinders University. Where these papers have been submitted for publication, the reference has been included below.

Chapter 3, which explores men's reactions to women's accusations of sexual harassment, has been submitted for publication as:

Barron, A. C., Thomas, E. F., & Woodyatt, L. (2022). MeToo, #MenToo: How Women's Accusations Shape Men's (Counter)Mobilization. Manuscript under review.

A Note About the Format of the Thesis

This thesis has been prepared as a series of papers to be submitted for publication. Chapter 3 of this thesis has been submitted for publication and is currently under review. Chapters 1 and 5 have been prepared in a traditional thesis format to give context to the thesis as a whole. Given the format of this thesis, the text within Chapter 3 is identical to the submitted paper. However, to ensure consistency across the thesis, I have made minor alterations to the numbering system of the studies, sections, tables, and figures within this paper. I have also replaced the use of first-person plural pronouns ("we", "our") with the singular "I" throughout. In addition, to avoid repetition, I have created one single reference list that can be found at the back of this thesis.

Chapter 1

Introduction

Online protests from disadvantaged groups have brought renewed attention to longstanding issues of inequality between groups. Movements such as #MeToo and Black Lives Matter have highlighted the pervasive nature of racial and gendered violence and led to increased societal discussions regarding ongoing discrimination against women and ethnic minorities. Notably, these movements have placed a spotlight on the role of the historically advantaged group in committing harm against disadvantaged group members. For example, the #MeToo movement has drawn attention to the over-representation of men among perpetrators of sexual harassment and assault. Similarly, the Black Lives Matter movement continues to demand federal law enforcement agencies acknowledge and address institutionalized racism and police brutality against Black people and members of ethnic minority groups.

The reactions of advantaged group members to movements advocating for social change are varied. While some choose to stand in solidarity with the disadvantaged group to challenge injustice (e.g., White people who attended Black Lives Matter protests across the US in 2020), others respond with backlash and resistance (e.g., men who claim that the #MeToo movement discriminates against their group; de Maricourt & Burrell, 2022; Lisnek et al., 2022). Indeed, both #MeToo and #BlackLivesMatter triggered reactionary counter-movements (#MenToo, #AllLivesMatter) that sought to advocate for the rights of the advantaged group (Becker, 2020; Boyle & Rathnayake, 2020; West et al., 2021).

Despite research on the importance of advantaged group allies in supporting disadvantaged groups to achieve social change (Drury & Kaiser, 2014; Subašić et al., 2008; 2018), little work has examined the factors that shape advantaged group members' reactions to

social movements that seek to change the status quo. What underlies advantaged group members' support for progressive versus reactionary forms of social change (that is, actions designed to "challenge or protect the status quo"; Becker, 2020, p.1)? How can we reduce backlash and mobilise advantaged group members in efforts to challenge inequality? The present thesis aims to address these questions in the context of men's reactions to the movement for gender equality. Specifically, I explore the social psychological factors that influence men's commitment to act for gender equality – as well as their *countermobilization* to feminist efforts in the form of actions designed to promote the rights of their own group.

In the current thesis I take the approach that the issue of men's support for (versus their resistance to) collective action for gender equality should be understood as a *dynamic intergroup phenomenon* (Drury & Reicher, 1999; 2009; Reicher, 1996; 2004; Stott & Drury, 1999). Drawing on research conducted within the social identity perspective (Bliuc et al., 2020; Lüders et al., 2022; Thomas et al., 2022a), I view online movements and counter-movements as a key site of intergroup conflict between (opposing) social groups. These groups can be grounded in shared ideologies or opinions (e.g., "I identify as a feminist/supporter of gender equality") – or relate to groups defined by existing status relations (i.e., "advantaged" or "disadvantaged" group identities; Jost et al., 2017).

I draw on insights from the needs-based model of reconciliation (Shnabel & Nadler, 2008) to suggest that the nature of men's opposition to the movement for gender equality may be shaped, in part, by their distinct identity-based needs as members of a structurally advantaged group. Specifically, I argue that men's support may hinge on the degree to which women's allegations of sexual violence threaten men's need for positive moral identity (Kende et al., 2020; Teixeira et al., 2020). I propose that men's need to defend their group's morality may not

only undermine their intentions to support the movement to end gender-based violence (e.g., support for #MeToo), but may also motivate their engagement in reactionary counter-movements designed to promote the rights of their own group (e.g., support for #MenToo).

Importantly, I suggest that movements for and against social change emerge and unfold dynamically over time as a product of both inter- *and* intragroup discussion and debate (Dixon et al., 2020; Louis et al., 2019; Thomas et al., 2022a; Vestergren et al., 2019). For example, in the context of the #MeToo movement, women's discussions with one another about their shared experiences of victimisation gave rise to their disclosures of sexual harassment and assault using the #MeToo hashtag. Women's disclosures, in turn, prompted men's debate, which led to the emergence of the counter-hashtag #NotAllMen. In reaction to #NotAllMen, women responded with #YesAllWomen. Drawing on the existing literature on the powerful effects of group discussion in shaping people's social identities (Postmes et al., 2005; Gee et al., 2007; Thomas & McGarty, 2009), I propose that targeting the nature of men's *intragroup* discussions regarding violence against women may provide a potential means of overcoming defensive reactions and increasing their commitment to act for gender equality at the intergroup level.

Although the empirical work presented in this thesis is conducted solely within the context of unequal status relations between men and women, my approach is underpinned by work on advantaged group identities more broadly, conducted within the frameworks of social identity theory (Tajfel & Turner, 1979) and the needs-based model of reconciliation (Shnabel & Nadler, 2008; 2015). I therefore refer to men as members of a historically advantaged – or "perpetrator" – group, and women as members of a historically disadvantaged – or "victim" group throughout this thesis (given that "victim" and "perpetrator" groups correspond in many

ways to advantaged and disadvantaged groups; Aydin et al., 2019; Kahalon et al., 2019; Siem et al., 2013; Shnabel et al., 2013).

In the following literature review I firstly provide a broad overview of the divergent ways that advantaged group members (e.g., men) can respond to efforts for progressive social change. Next, I outline how modern social movements from disadvantaged groups typically call into question the moral character of the advantaged group, leading those highly invested in their group membership to experience a threat to their moral identity. I discuss how advantaged group members' concerns regarding their group's morality can motivate defensive reactions, before proposing that defensiveness should reduce men's support for gender equality and may also underlie their counter-action to promote the rights of their own group. Finally, I discuss a potential means of mitigating defensiveness and engaging advantaged group members in solidarity-based action for the disadvantaged group in light of evidence suggesting the importance of opinion-based social identities in creating new avenues for people's commitment to social change.

Advantaged group members' reactions to inequality

In the collective action literature, people's responses to inequality have been broadly divided into actions that seek to challenge the existing status quo and reduce inequality between groups – or actions designed to *oppose* social change and uphold the existing status hierarchy (Jost et al., 2017). These distinct forms of action have been referred to as *progressive* (or *system-challenging*) collective action, and *reactionary* (or *system-supporting*) collective action, respectively (Becker, 2020; Choma et al., 2020; Jost et al., 2017; Osborne et al., 2019). System-challenging action often includes the support of advantaged group members, who act "in solidarity" or "allyship" with disadvantaged group members to achieve social change

(Droogendyk et al., 2016; Iyer & Leach, 2010; Louis, 2009; Subašić et al., 2008; 2018). Prominent examples of advantaged group allies include heterosexual individuals who act in support of LGBTIQA+ rights (Russell, 2011), White people who advocate for justice for Black and Indigenous people (Clark, 2019), and men who support the fight for women's rights (Drury & Kaiser, 2014; Hardacre & Subašić, 2018).

Conversely, system-supporting collective action often involves advantaged group members' attempts to *oppose* equality and defend the prevailing status hierarchy that privileges their group (e.g., the rise of populist "alt-right" or "far-right" movements; Becker, 2020; Mikołajczak et al., 2022a; Selvanathan et al., 2020b). Of course, it should be noted that it is also often the case that advantaged group members opt to do nothing in the face of injustice – that is, they disengage entirely from debates surrounding the rights of the disadvantaged group (captured by work on *inaction*; see Elad-Strenger et al., 2022; Stroebe et al., 2019).

What predicts advantaged group members' support versus resistance: The role of group-based threats

Social psychological research on advantaged group members' support and opposition to social change has focused predominantly on how the increasing rights of disadvantaged group members can lead members of the advantaged group to feel that their position in the status hierarchy is under threat (Brown et al., 2022; Craig & Richeson, 2014; Craig et al., 2018; Domen et al., 2022; Dover et al., 2016; Reicher & Ulusahin, 2020; River-Rodriguez, 2022; Shepherd et al., 2018). Advantaged group members' perception that they are competing with the disadvantaged group for power and resources has been associated with their reduced support for progressive policies designed to protect minority group members from prejudice and discrimination (Craig & Richeson, 2014; Leach et al., 2007b; Norton & Sommers, 2011). In a

similar vein, research has linked conservative ideologies such as right-wing authoritarianism (RWA) and social dominance orientation (SDO) with advantaged group members' opposition to system-challenging action (e.g., Black Lives Matter), and their engagement in system-supporting action (e.g., White nationalist movements that seek to protect "White power"; Choma et al., 2020; Burley & Ross, 2019; Holt et al., 2022; Selvanathan et al., 2020b).

However, movements for progressive social change are not only concerned with the disadvantaged groups access to rights and resources. Protest from disadvantaged groups also highlight the unearned privileges and immoral actions of the advantaged group (Kende et al., 2020; Okuyan & Vollhardt, 2022; Shuman et al., 2022; Teixeira et al., 2020). Indeed, commentary surrounding the #MeToo movement notes that while #MeToo was instrumental in emphasizing the need for structural, legislative change to prevent violence against women, it also brought into sharp relief men's over-representation as the perpetrators of sexual violence (de Maricourt & Burrell, 2022; Hill, 2021; Lisnek et al., 2022). Hill (2021) discusses how, as #MeToo gained traction online in 2017, it increasingly became an "accountability" movement concerned with promoting justice for victims and retribution for male perpetrators (p. 10). In this way, movements like #MeToo not only challenge existing power relations between men and women, but also call into question men's moral character.

A significant body of research in social psychology has shown that members of advantaged groups are particularly sensitive to information suggesting their group has done wrong (Branscombe et al., 1999; Kahalon et al., 2019; Knowles et al., 2014; Sullivan et al., 2012). According to the needs-based model of reconciliation (Shnabel & Nadler, 2008), disadvantaged group members – who are often the victims of discrimination – can experience a threat to their need for power and agency. Conversely, advantaged group members – who are

often accused of being prejudiced against the disadvantaged group – experience a heightened need for morality and acceptance (Nadler & Shnabel, 2017; Shnabel et al., 2009). Research suggests that this is particularly the case for advantaged group members who are highly invested in their group membership (Branscombe et al., 1999; Iyer & Ryan, 2009; Leach et al., 2007a; Tajfel & Turner, 1979).

Drawing on the framework provided by the needs-based model, research has shown that advantaged group members' support for social change is influenced by their perception that their group's moral image is under attack (see Kahalon et al., 2019; Kende et al., 2020; Teixeira et al., 2020). For example, Kende et al. (2020) showed that men's need to defend their group's moral reputation was associated with their reduced support for the #MeToo campaign. However, to date, this work has focused on how morality threats may undermine advantaged group members' *support* for progressive social change, without considering how concerns about their group's morality may *also* mobilise advantaged group members to advocate for the rights of their *own* (privileged) group.

In the current thesis I argue that morality concerns (and group members' subsequent defensiveness) may play an important role in shaping *both* advantaged group members' support for the disadvantaged groups' protest, *as well as* their (counter)mobilization to promote the rights of their own advantaged group. Indeed, in recent years, online counter-movements from advantaged group members (#NotAllMen, #MenToo, #AllLivesMatter) have often been characterized by defensive attempts to protect their group's moral image (e.g., by denying their group's role in perpetrating harm; Bilali, 2013); minimizing the severity of the wrongdoing (Bilali et al., 2012; Leidner et al. 2010); blaming the victim(s) (Bandura, 1996; 1999), and/or arguing that their group has suffered *more* than the disadvantaged group (*competitive*

victimhood; Noor et al. 2012; Sullivan et al., 2012; Young & Sullivan, 2016). In the context of responses to the #MeToo movement, men's backlash included strategies designed to downplay (or outright deny) the issue of violence against women. #NotAllMen was used to claim that sexual violence is only perpetrated by a few "bad apples" (thereby allowing men to deny the structural nature of gender inequality by positioning it as a problem that is only attributable to a "deviant" few; Flood, 2019). #MenToo asserted that men are *also* the victims of sexual harassment and violence. As victims are often viewed as morally superior to perpetrators, claims to victimhood (such as "#MenToo") function to defend against moral image threats by asserting one's own group has "moral credentials" (Sullivan et al., p. 102; see also Noor et al., 2012; Young & Sullivan, 2016).

Defensive reactions to reminders of ingroup harm have received substantial attention in social psychology (see Bilali & Vollhardt, 2019; Demirdağ et al., 2019, Kahalon et al., 2018; Noor et al., 2012; Sullivan et al. 2012). However, to date, little work has examined whether defensiveness may motivate members of advantaged groups to engage in reactionary forms of social change (that is, movements designed to promote the rights of their own privileged group). Accordingly, I propose that defensiveness – that is, the various strategies people employ to protect against threats to their personal or group identity – may underlie men's countermobilization to feminist movements as a means of deflecting attention away from the (morally threatening) issue of men's violence against women (Chapter 3). Specifically, I explore the relationship between men's social identification (as members of a structurally advantaged group responsible for perpetrating harm) and their concerns about their group's moral image. I suggest that the strength of this relationship will vary based on the extent to which women's claims implicate men as the perpetrators of sexual harassment. Men's concerns about their

group's morality should motivate their engagement in defensive strategies to reduce this threat. I test whether defensiveness, in turn, is associated with men's *reduced* support for actions designed to promote women's rights, and *increased* support for counter-action intended to advance the rights of their own group.

Overcoming defensiveness and engendering men's support for gender equality

Men's participation in collective action for gender equality requires acknowledgement of their group's power, privilege, and history of oppression over women. However, the research reviewed so far suggests that this may be particularly problematic, as members of advantaged groups are strongly motivated to defend against information suggesting their group has done wrong (Cohen, 2001; Bandura, 1996; Bilali & Vollhardt, 2019; Branscombe et al., 1999; Doosje et al., 1998; Peetz et al., 2010). Importantly, research suggests that defensive reactions to social change are likely to emerge to the extent that advantaged group members (e.g., men) are defining themselves in terms of a group responsible for perpetrating harm – that is, they view their advantaged group membership (i.e., their social identification as a man) as a key part of their self-concept (Branscombe et al., 1999; Branscombe & Miron, 2004).

However, as stated earlier, advantaged group members *do* act in solidarity with members of the disadvantaged group to challenge injustice and inequality (Iyer & Ryan, 2009). Research suggests that advantaged group members' support for progressive social change can be explained by their social identification with groups grounded in concrete *opinions* about the world (rather than pre-existing social categories based on race, gender, or sexual orientation; Bliuc et al., 2007; McGarty et al., 2009). Opinion-based groups (e.g., "I identify as a supporter of gender equality") cut across pre-existing groups defined by "advantage" or "disadvantage" – and in this way capture how advantaged group members can come to act in support for equality, as well as how

members of disadvantaged groups participate in actions that seemingly go against their group interests; see Mikołajczak et al., 2022). This work suggests that a potential avenue to overcoming men's defensiveness (and engendering men's support for equality) is to facilitate their identification with groups based in acknowledging and addressing the issue of gender inequality. The literature on *how* people come to identify with opinion-based groups offers key insights into how this may be achieved.

Group interaction as a facilitator of opinion-based identities for equality

Research examining the processes through which people come to participate in coordinated, collective behaviour suggests that opinion-based identities rooted in support for social change can form through an iterative process of communication and interaction between people (Bongiorno et al., 2016; Postmes et al., 2005, Thomas & McGarty, 2009, Smith & Postmes, 2009; 2011). Through interacting with like-minded others, people can come to an understanding that their grievances and opinions about the world are shared (Smith et al., 2015). Importantly, in discussing *how* to address a particular social or political problem through joint action, people can come to inductively generate social norms that value coordinated collective action. It is through the internalization of these norms – and the understanding that these norms are shared by others – that new opinion-based identities grounded in support for action are crystallized (Thomas et al., 2009).

The role of group interaction in facilitating opinion-based identities predicated on supporting prosocial (versus antisocial) action has received empirical support across a number of intergroup contexts (e.g., Thomas & McGarty, 2009; Thomas et al., 2016; 2019; Smith & Postmes, 2009; 2011). For example, Thomas & McGarty (2009) found that participants who engaged in group interaction regarding strategies to address poverty in third-world countries

(relative to those who completed the same task in isolation) reported greater intentions to act to address poverty, an effect mediated by their social identification as a supporter of efforts to achieve clean drinking water for people in developing nations. Conversely, in the context of attitudes towards immigration in the United Kingdom, Smith & Postmes (2011) demonstrated that group interaction gave rise to *hostile* intergroup behaviour (casting a vote for a fictitious anti-immigration political party) when participants discussed harmful stereotypes about immigrants in Britain.

Taken together, these findings highlight the influential role of group interaction in shaping social identities based in support for either prosocial (or antisocial) intergroup behaviour. However, to date, research has not tested whether group interaction could be strategically used as a means of fostering men's commitment to act for gender equality. Accordingly, in the present thesis I suggest that targeting the nature of men's intragroup interactions could provide a productive means of restructuring problematic intergroup boundaries that position men's role in the fight for gender equality as either bystanders or perpetrators (Subašić et al., 2018; p., 709). That is, I test whether engaging men in group interaction about addressing sexual harassment and assault against women can facilitate the development of an identity grounded in *acknowledging* – rather than denying – the issue of violence against women (Chapter 4). I explore whether this identity reduces men's defensiveness regarding the issue of gendered violence, and, in turn, increases their commitment to act to combat gender inequality.

Summary and Overview

The present thesis is guided by two key questions: (1) what underlies advantaged group members support for progressive versus reactionary forms of collective action? and (2) how can we reduce backlash and mobilise advantaged group members in efforts to challenge inequality? I

examine these two questions in the context of men's support for (versus their resistance to) collective action for gender equality. I take a social identity perspective and argue that the nature of men's opposition to the movement for gender equality may be shaped by their unique identity-based needs as members of a structurally advantaged group accused of perpetrating harm (Shnabel & Nadler, 2008; 2015). I further argue that opinion-based identities grounded in support for addressing gender inequality may be particularly well placed to overcome men's resistance to feminist efforts and, in turn, foster their commitment to act for gender equality.

I begin this thesis by providing an overview of the nature of online movements for and against progressive social change in the context of the #MeToo movement within Australia (Chapter 2). Using a structural topic modelling approach to analyse Twitter data, I explore how Twitter users expressed support or opposition to the March 4 Justice movement to end gender-based violence. In this chapter I argue that social media platforms are a key site of intergroup discussion and debate regarding contentious social and political issues. In particular, I show that supporters of March 4 Justice used social media to air grievances, express solidarity, and coordinate offline actions. I contrast the tactics used by supporters with evidence that Twitter users' resistance to March 4 Justice was largely characterized by concerns about the impact of sexual assault allegations on men's moral and social reputation, and the use of defensive strategies that sought to downplay (or outright deny) the issue of violence against women.

Given the centrality of concerns about men's moral image in tweets surrounding the March 4 Justice movement (Chapter 2), in Chapter 3 I explore the influence of men's social identification (as members of a structurally advantaged group responsible for perpetrating harm) and women's accusations of victimisation on men's need for morality, defensiveness, and collective action intentions (both system-supporting and system-challenging action). I report on

three studies showing that men's social identification is associated with their reduced intentions to act for women's rights, and increased intentions to advance the rights of their own group. These effects were mediated by men's need for morality and defensiveness. The findings of this chapter suggest that defensiveness may not only present a barrier to men's engagement in solidarity-based action for gender equality, but that it may also motivate their participation in action designed to promote the rights of their own (privileged) group (as a form of reactionary action). Chapter 3 anticipates an important point of this thesis: Namely, that to understand the success of disadvantaged groups' social movements – and to engender advantaged group members as allies – we must understand how advantaged group members perceive of and react to such movements. Given that social media is a key site at which advantage and disadvantage, privilege and oppression are debated, I argue that online interactions provide accessible test beds for developing our understanding of these issues.

In the fourth chapter of this thesis I explore a method designed to mitigate men's defensiveness regarding the issue of gendered violence and increase their engagement in collective action for gender equality. This chapter draws on a group-discussion paradigm that has been found to foster the development of *opinion-based* identities grounded in support for prosocial collective action via engaging participants in interaction with one another. Accordingly, I propose that group discussion may be fruitful means of reducing defensive reactions (and increasing men's engagement in action for gender equality) via the formation of an identity grounded in taking collective action to address inequality between men and women. I show that group interaction boosts men's commitment to act for gender equality via increased identification as a supporter of efforts to address gendered violence and reduced defensiveness regarding the issue of violence against women (Study 5). Importantly, the effectiveness of group

interaction on men's commitment to act for gender equality is contingent on the content of their discussions with one another (Studies 7 and 8). When men's discussions centred practical collective action strategies to address the issue of sexual harassment and assault against women (relative to discussions that solely focused on the *issue* of gendered violence), this increased their identification as a supporter of addressing sexual harassment and assault against women. Identification, in turn, was positively associated with men's intentions to act for gender equality – as well as a behavioural measure of social action: clicking on a link relevant to expressing support for political, legislative change designed to protect survivors of domestic and sexual violence and their families. The findings of this chapter provide an important basis for further investigation into the individual and contextual factors that are likely to influence the effects of group interaction on men's commitment to act for gender equality.

In Chapter 5 I summarise the key findings of this thesis and consider the theoretical and practical implications regarding the factors that may promote (or hinder) men's commitment to act for gender equality. I also discuss empirical issues related to the study of the dynamic and interdependent processes that underlie people's support for progressive (versus reactionary) forms of collective action. Finally, I consider the limitations of the current thesis and offer directions for future research to further our understanding of advantaged group members' support for equality.

Chapter 2

Exploring the Nature of Online Support and Resistance to Gender Equality:

A Case Study of Responses to #March4Justice on Twitter

Social networking sites (SNS) are a key platform for modern social and political protest and intra- and intergroup debate regarding issues of inequality between groups (Segerberg & Bennett, 2011; Greijdanus et al., 2020; Gomez & Kaiser; 2019). As members of disadvantaged groups become increasingly empowered to use social media to bring awareness to their experiences of victimisation, reactionary counter-movements from those seeking to defend the status quo have become commonplace (Hodson et al., 2022; Selvanathan et al., 2020; van Haperen et al., 2022). Research has begun to use computational approaches to explore the nature of online social movements for and against social change. For example, studies have examined the language use of people advocating for the rights of the disadvantaged group (see Foster & Rathlin, 2022; Nau et al., 2022; Roth-Cohen, 2021) and extreme forms of hate speech within online fringe communities (see Baele et al., 2021; Dickel & Evolvi, 2022). However, this work has tended to investigate support for progressive (versus reactionary) social movements as though they occur independently from one another. To date, less work has focused on how social media provides an important site of contestation *between* groups that seek to change the existing social system and those that act to defend it.

I argue that examining the dynamic and interdependent nature of social movements and counter-movements (that is, the way they are shaped by interactions both within *and* between groups) is critical to understanding the dialectical nature of collective action for social change (Dixon et al., 2020; Drury & Reicher, 2000; Thomas et al., 2016; 2022a; Vestergren et al., 2019). The purpose of the present chapter is to therefore provide an exploration of the nature of online

support *and* backlash towards the movement to end gender-based violence within Australia. To this end, I analyse social media data sourced from Twitter during a resurgence of the #MeToo movement in Australia in early 2021, following a series of sexual assault allegations made against high profile members of Australian politics. On March 15, 2021, over 110,000 people participated in organised "March 4 Justice" protests across Australian capital cities to call for an end to sexual assault, harassment, and violence against women. I collected and analysed a unique data base of tweets posted between this time (January 1st – June 30th, 2021) containing the terms "#MeToo + #Auspol" or "#March4Justice". I focus on this particular time period as it reflects people's initial reactions to the allegations of sexual assault and harassment emerging from Australian parliament, as well as capturing the lead up to (and period shortly after) the March 4 Justice rallies.

Using a corpus of 152, 096 tweets, I use a structural topic modelling method (STM; Roberts, 2014; 2019) to identify, organize, and explore themes (or "topics") in the tweets surrounding the March 4 Justice movement. Structural Topic Modelling is a common machine learning technique that works on the assumption that a collection of texts (e.g., tweets) contains several latent topics, and that those topics are made up of a mixture of words (Blei et al., 2003; Griffiths & Steyvers, 2004; Robinson & Silge, 2017). Topics are therefore identified based on the co-occurrence of key words in a corpus of text (Roberts et al., 2014).

It is important to recognise that topic modelling is a data-driven method. I use the results of the structural topic model to firstly explore how tweets are broadly clustered together – that is, to detect the overarching topics of Twitter discussions surrounding the March 4 Justice campaign. I then take a more fine-grained approach by qualitatively reading the top tweets from each topic to analyse these clusters in more detail. In combining these two approaches (data-

driven and theoretical), I aim to provide an overview of the nature of online support and resistance to the movement for gender equality within Australia in 2021.

The Current Context

Early 2021 saw a resurgence of the #MeToo movement in Australia following a number of allegations of sexual assault and misconduct against women involving current and former federal politicians. Shortly after Grace Tame was named Australian of the Year for her advocacy regarding the impacts of sexual violence, former Liberal party staffer Brittany Higgins spoke out publicly about her experience of sexual violence in a ministerial office within Parliament House. In March 2021, historical sexual assault allegations were made against former Australian Attorney-General Christian Porter. In the months following, further allegations of sexual harassment and misconduct emerged from Parliament House, spurring national conversations about a culture of misogyny, discrimination, and violence against women (within both federal politics and the broader country).

The culmination of these events set the stage for the March 4 Justice rallies, a series of organised protests that took place in capital cities and towns across Australia on March 15, 2021. Organisers of March 4 Justice outlined four key demands: (1) independent investigations into all cases of gendered violence; (2) full implementation of all 55 recommendations from the *Respect@Work: Sexual Harassment National Inquiry* report; (3) increased funding for gendered violence prevention to world's best practice standards; and (4) the enactment of a federal Gender Equality Act to promote gender equality, including a gender equity audit of parliamentary practices (Lee, 2021).

Study 1

Data and Methods

Data collection

I constructed a corpus of 152, 096 tweets using the R library package *academictwitteR* to access Twitter's API v2 full-archive endpoint through Twitter's academic research program (Twitter, 2022). First, I searched for 30, 000 tweets posted in English between January 1st and July 30th, 2021, that included the terms "#MeToo" AND "#Auspol" OR "#March4Justice". Boolean operators "AND" and "OR" meant that the search was restricted to tweets discussing gendered violence within the same (Australian) context. These operators also reduced the likelihood of collecting tweets that included the phrase "#Metoo" but were unrelated to sexual violence (given the hashtag is commonly used on Twitter to indicate agreement). The search excluded retweets in an attempt to reduce duplicates at the point of data collection. This initial search resulted in 19, 027 tweets, of which 18, 500 remained following the removal of duplicates.

Following this, I collected a subset of the replies to the initial 18, 500 tweets returned by the initial key word search. Given backlash on Twitter often emerges in people's replies to prominent tweets, by collecting both tweets *and* their replies I hoped to better capture debates between users (and thus increased variation in supporters and opponents). Collecting the replies from a subset of 12, 545 tweets from the original 18, 500 yielded a total of 597, 793 tweets. This corpus was subsequently reduced to 152, 096 tweets after the removal of duplicates.

Data Cleaning

Data were pre-processed in R using tidy text principles (Robinson & Silge, 2017). To create a document-term matrix to be analysed using a Structural Topic Model, tweets were

cleaned following standard pre-processing procedures: the removal of hyperlinks, emojis, numbers, punctuation, and common "stop words" (i.e., articles and pronouns). I further removed terms that were too infrequent (appearing in less than 0.01% of the documents) to avoid generalization and prevent overfitting the model (as per the approach outlined by Robsinson & Silge, 2017).

Analytical approach: Structural Topic Modeling

Topic modeling is a common text mining approach used to organize and explore large amounts of data into given themes or "topics" (Blei et al., 2003; Griffiths & Steyvers, 2004; Robinson & Silge, 2017). In the current chapter I used a Structural Topic Model (STM, Roberts et al., 2019) to extract topics from my corpus of 152, 096 tweets. STM is an unsupervised topic modeling technique that is used to describe relationships among words that co-occur together within documents (i.e., within a corpus of tweets). Similar to other commonly used topic modelling approaches such as Latent Dirichlet Allocation (LDA; Blei et al., 2003), STM works by inferring latent topics based on the co-occurrence of key words which are assumed to be semantically related (Roberts et al., 2014).

Model Assessment. A well-documented limitation of topic models is their sensitivity to parameters specified by the researcher – namely, how many topics (k) to estimate in a given corpus (Steyvers & Griffiths, 2007). I therefore used R's *ldatuning* package to estimate the optimal number of topics to be used in the model, based on previously established metrics commonly used to assess topic models (Nikita, 2015). This assessment calculates four different metrics ("Griffiths2004", "Arun2010", "CaoJuan2009" and Deveaud2014") to estimate model fit at k = 2-25 topics (Arun et al., 2010; Cao et al., 2009; Deveaud et al., 2014; Griffiths & Steyvers, 2004). The goal is to find extremum (that is, the maximum or minimum values) on a scale from

0-1, such that Griffiths2004 and Deveaud2014 are maximized and Arun2010 and CaoJuan2009 are minimized. Results (see Figure 1) indicate that an optimal number of topics was 14. Figure 1 shows that when $k = 14$, both Griffiths2004 and Deveaud2014 were between 0.50-0.80, while both Arun2010 and CaoJuan2009 were below 0.50 (Arun2010 is below 0.25). I therefore estimated a topic model with 14 topics.
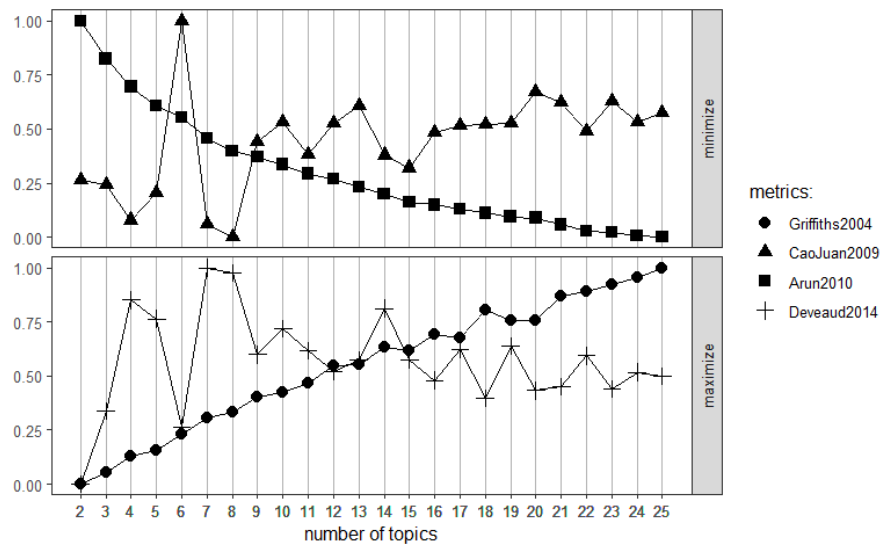
Figure 1. Plot of Optimal Number of Topics (*k*) to be Used in the Structural Topic Model Against LDA Metrics.

Results

One topic model with 14 topics, 152, 096 documents and a 54, 935-word dictionary was generated using R's *stm* package (Roberts, 2019). Figure 2 shows the most prevalent topics across the corpus, as well as the seven most-frequent terms associated with each topic. As can be seen in Figure 2, in general, each topic is equally distributed across the corpus. Table 1 includes a label for each topic, each topic's top terms (the highest probability words in a topic), the most frequent and exclusive terms per topic (FREX terms: FRequency and EXclusivity), the

prevalence of each topic across the corpus, and tweets highly associated with the particular topic. Topics were labelled based on a qualitative reading of the top terms, FREX terms, and top tweets corresponding to each topic. By extracting the tweets highly associated with each topic I allowed for a more meaningful and comprehensive understanding of the data.
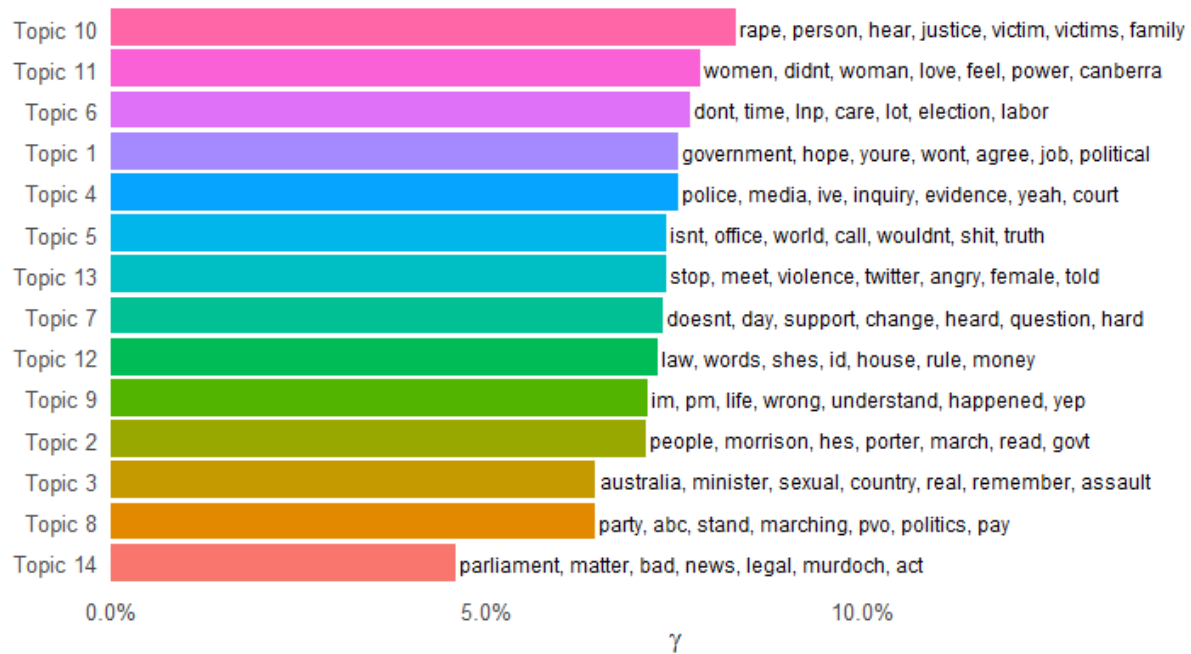
Figure 2. Topic Prevalence in the #March4Justice Corpus with the Highest Probability Words for each Topic.

Table 1. Structural Topic Model for Tweets from the #March4Justice Corpus. Table Includes Topic Labels, the Top Terms Per Topic,

the Most Exclusive Terms Among Topics (FREX terms), and Tweets Highly Associated with Each Topic.

No.	Topic label	Highest probability terms	FREX terms	Exemplar tweets	%
1	Criticism of the government	government, hope, youre, wont, agree, job, political, true, australian, speak	thankyou, fabulous, nah, xx, perth, haha, cool	Sacked by the Howard Government. How fracking useless would you need to be to be sacked by "the most profligate Australian government ever"? You're saying the Liberal Government honoured left wing policies? Where? Which legislations were continued? It was legislated against the Lib governments wishes and it was repealed by a Lib government. How is that being honoured? POV is on the money...a Frat Boy and a useless twat that should have stuck to tennis...LNP down the fcuking gurgler	7.54%
2	Comparison of political leaders	people, morrison, hes, porter, march, read, govt, christian, hold, scott	ides, onselen, military, van, scott, junta, gun	Bob Hawke took to the podium before angry farmers... John Howard did the same in front of angry gun owners in a bulletproof vest... Let's see if Morrison is up to the task. Isn't it funny how history repeats itself ~ "Beware the ides of March". But instead of knives in Cesears back, its the metaphorical knives 2 the throat of Morrison's evil government wielded by those who shall never remain silent again over the tyranny of this depraved government Hawke found the time to go outside. John Howard faced a crowd of angry men with guns (albeit wearing a ballistic vest). Scott Morrison can't even face a crowd of women. Scott is scared of women.	7.10%
3	Desire for political action	australia, minister, sexual, country, real, remember, assault, live, action, girls	rats, sewer, rat, fu, ho, prime, presumption	Having delivered the ultimate FU to the women and girls of Australia, may I suggest we women and girls deliver an equally vehement FU to the LNP as a whole at any and all elections for the next 5 decades? They may then learn out worth. Wanna join us?	6.45%

(2) .... Sally McManus, Michele O'Neil, Animata Conteh-Biger, Biff Ward, Virginia Hausegger, Avan Daruwalla, and Maddie Chia. "Women have the strength, resilience and ability to transform the world, and we must do it, with love, compassion, empathy and harmony.

Need independent inquiry into AC sexual assaults & mandatory repting of assault/abuse in AC, as with child care.

4	Criticism of criminal justice system	police, media, ive, inquiry, evidence, yeah, court, story, leave, nsw	dicks, nsw, doors, swinging, hendry, sa, coroner	Is 4C asking us to read between lines? The longer NSW Police (any Police) delayed process of signed statement into 'Sworn Affidavit' so it's entered into evidence of activated proceedings, the longer the delay in bringing it to court, and critically, increasing claimant duress

Katharine first reported this to SA Police in Nov 2019. It took NSW Police until Feb 2020 to contact her. It then took another 4 months despite investigators contacting her on at least five occasions to arrange a formal statement, which she later declined on June 23rd.

No it isn't. The Coronial Inquest is yet to be completed and a Judicial Inquiry could still be initiated. Also what needs investigating is why did the NSW police take so long to get a statement. The covid excuse doesn't stack up. | 7.54% |
| 5 | Anger at government | isnt, office, world, call, wouldnt, shit, truth | photo, op, ops, agreed, cheers, truth, yup | Private meeting means he sets the parameters with his own private photographer, thus a PR stunt. Nothing in discussion with him in his private office can be made public. His own spin will deny anything negative.

This corrupt Government wouldn't know the truth if it hit them in the face....just more spin and BS...we've had enough..

If anyone goes into his fucking palace just to get a photo op for him with his photographer and only to get his fucking face everywhere, I'm not marching. He has to come out of his lair like a normal human or fuck that shit | 7.39% |

6	Impact on federal election	dont, time, lnp, care, lot, election, labor, ill, vote, theyre	vote, election, marise, labor, wa, getup, karma	This train wreck, the LNP, is just proving what we all knew. They don't deliver on promises, they are out of touch with contemporary women's issues and they don't care about the environment AT ALL! Mass extinctions... no we need the LNP to become extinct! Labor wins in a landslide in WA, the federal LNP lurching between scandal and crisis and the next federal election less than twelve months away. Not enough time to turn their situation around but definitely enough time for the Murdoch press to lose any remaining credibility This Court action will not be finished before the next Fed. Election. The LNP have no chance next Election if this action has not been resolved i can see a WA Federal Election type result, Labor will smash them then comes the Federal ICAC with enormous Powers.	7.70%
7	Desire for social change	doesnt, day, support, change, heard, question, hard, issue, womens, respect	lol, conquer, cow, omg, retweet, voices, lying	Zero problem here. My support is unconditional. I support the movement because it's right and just and loooooong overdue. Kia kaha. Go. Change is odd - for years nothing seems to happen despite insufferable wrongs and then there is a spark and the whole rotten edifice goes up in flames. This is such a moment. To borrow from W B Yeats: "All changed, changed utterly: A terrible beauty is born" Join your union, form work place consulting groups, write to your representatives, put yourself up for leadership roles. Hit the activation button - enough of this, now its our time!	7.34%
8	Expressions of solidarity	party, abc, stand, marching, pvo, politics, pay, powerful, speech, line	blah, grace, tames, digging, sociopaths, cheat, penny	If Porter settles for other than unconditional apology, plus damages and costs he loses. If ABC apologises unconditionally, they have to sack the news editor, producer of and Milligan. Breakfast News have probably had their orders. Ms Tame's speech was too powerful to be forgotten, just as Julia Gillard's misogyny speech has never been forgotten.	6.43%

I hope so..Brittany & Grace made powerful brave inspirational speeches today..this is not a war against men..it's not a partisan war..we need good men (Sam Neill) to stand with us..for their daughters, partners, sisters, nieces, mothers, aunts...

9	Anger at former Australian Prime Minister Scott Morrison	im, pm, life, wrong, understand, happened, yep, bit, hell, makes	ha, yep, kathy, melbourne, macro, hell, sherriff	How is the PM not bullying? How is the PM not degrading an innocent female. How can we believe a PM that says he will fix the problem when he plainly does it himself without any remorse or remedy. Sorry isn't the hardest word... silence is. I'm sorry for what happened to you that one night destroyed your whole life and the PM thinks it's a side issue The man-child knows he is doing the wrong thing and probably has residual feelings that maybe he should be ashamed of his actions. I bet his upbringing reinforced his feelings of shame.	7.13%
10	Impact of sexual assault allegations	Highest Probability: rape, person, hear, justice, victim, victims, family, alleged, allegations, absolutely	higgins, allegation, alleged, brilliant, victims, brittany, nope	The two cases related to Parliament have not been proven! One allegedly occurred over 30 years ago. The alleged victim apparently said she had dinner with her alleged rapist 6 years after the alleged incident. The other case the staffer decided not to lodge a formal complaint. You can say what you want but we all know it is false accusers that make it harder for genuine victims to get justice. False accusers and the people that support them instead of deterring them. where is the deterrent to false accusations of rape today??? This is not because of any political dictum like "Believe women." It's because the deceased alleged victim's story looks exactly like tens of thousands of date rapes that happen every year, and nothing at all like a false rape accusation!	8.13%
11	Identity concerns	women, didnt, woman, love, feel, power, canberra, liberal, white, male	spirit, helen, solidarity, cower, reddy, grateful, love	Gee ..Great time to be a White Male. Did or do all these women really hate their fathers and husbands that much? Oh. it is not all males then ? Just a few bad apples ? 99 % of men do not rape or bash women , please stop tarring all men with the same brush. Adjust your demands .	7.83%

Stop using such a tragic incident to attack men. Not all women are innocent harmless beings. Some men are fearful of women too and are also attacked on our streets.

No one here is being racist... We're discussing intersectional feminism. White women, indigenous women, black women, asian women, all experience misogyny, but only women of colour experience racialised misogyny. Brushing that under the carpet as "we're all women" is not helpful

12	Instrumental use of social media	law, words, shes, id, house, rule, money, happy, talking, coming	rule, city, deaf, tone, birthday, updates, spot	Transport Canberra will be running free shuttle services from London Ct (outside Mooseheads) from 10.30am to Parliament House for those travelling to today's Women's March4Justice. Rtrn shuttles will depart from behind Old Parliament House from 1.30pm back to the City interchange	7.27%

Seems like 'rule of law' was always plural. Very, very plural. And institutionally, systemically, riddled with prejudice (there's a hint: pre-judice).

Michael O'Brien has declared he has no faith in the DHHS or its staff. Everyone is a "minion" of Daniel Andrews. DHHS will be restructured after Michael O'Brien is elected premier in November 2022. No current DHHS staff will be re-employed.

13	Concerns about racial violence	stop, meet, violence, twitter, angry, female, told, abuse, home, protest	language, fantastic, yay, gendered, sign, space, volumes	Lets not 4get the women in our prisons. Aboriginal TSI women experience family violence at a higher rate than the broader Aust community, that the majority of Aboriginal and TSI women in prison have experienced physical or sexual abuse.	7.38%

I once saw an exhibition of the clothing women/girls were wearing when they were raped. One outfit was a colourful striped skivvy, denim dungarees and sneakers, belonging to an 8yo girl Katie.

No, I'm not. Petition is 4 indept investgs into all cases of gendered violence & timely referrals 2 appro auths, full pub acc 4 findings, implement all recoms in AHRC report, lift public $ 4 gendered violence prevention 2 world's best prac & enact a federal Gender Equality Act.

| 14 | Economy and media | parliament, matter, bad, news, legal, murdoch, act, politicians, australians, free | tax, murdoch, propaganda, paying, wiped, jenny, andrews | Murdoch and real news ... lol ;) .... from the company that says ... opinion and commentary is cheaper and more profitable than news :)

If the "citizen" normally pays NO TAX on any other income, then no tax would be paid on franked dividends either.

basic economics every Australian needs to know. 1: "subsidies" is an increase in Taxation. 2: EVERYTIME Govt interferes in the "free" market, prices RISE. 3. Subsidies are the MAIN reason for growing Wealth Inequality. 4. Govt interference in the economy is the PROBLEM. | 4.59% |

What is the nature of online support for the March 4 Justice movement?

The results of the STM show that support for the March 4 Justice movement was characterized by (1) expressions of anger/perceptions of injustice, (2) expressions of solidarity, (3) demands for social change, (4) the instrumental use of Twitter, and (5) debates regarding the inclusivity of "#MeToo".

Expressions of anger/perceptions of injustice

Several topics identified by the STM reflect users' anger and sense of injustice about the issue of gender-based violence. These topics capture anger directed at the government's perceived inaction regarding the issue of violence against women (T5), the response of the criminal justice system to cases involving sexual abuse (T4), and the behaviour of former Australian Prime Minister Scott Morrison (captured in Topics 2, 5, and 9). Closer inspection of the top tweets associated with these topics show users expressing moral outrage at Morrison's decision not to attend the March 4 Justice protest held outside of Parliament House in Canberra on March 15, 2021 (T2). Tweets in Topic 2 contrast Scott Morrison's behaviour unfavourably with that of former Australian Prime Ministers, Bob Hawke and John Howard, in order to criticise his response to the issue of gender inequality (T2). Topic 2 also includes discussions about the allegations of sexual assault made against former Attorney-General Christian Porter. Interestingly, this topic also includes tweets from users raising awareness of crimes against humanity occurring in Myanmar, following the February 2021 coup that saw Myanmar's military junta seize power (Charney, 2021). This particular finding highlights the heterogeneity of tweets across the corpus (and of online social movements more generally) and suggests that social media users may strategically reply to tweets about a particular social issue in an attempt to co-opt support for other social or political causes.

Topics 5 and 9 reflect criticism levelled at the Liberal Party of Australia (LNP). The Liberal Party is one of the major political parties in Australia and was in power during the

period leading up to and directly after the March 4 Justice protests. Top terms and tweets associated with these topics suggest that former Liberal Party leader Scott Morrison was the target of people's anger. Similar to Topic 2, tweets in Topics 5 and 9 express anger about Morrison's offer to meet the organisers of the March 4 Justice protests in private, rather than in-person outside of Parliament House. Topics reflecting anger at the government's perceived silence regarding the issue of sexual violence (both within federal parliament and the wider country) seemed to spur broader criticisms of the Liberal party (captured in Topic 1), discussions regarding the impact of the March 4 Justice movement on the outcomes of the next Australian federal election (Topic 6), and discussions regarding the Australian media coverage of the allegations surrounding federal parliament (Topic 14).

Evidence of anger in the March 4 Justice corpus is consistent with social psychological research that has long established anger as a predictor of people's engagement in social and political protest (Thomas & McGarty, 2009; Simon & Klandermans, 2001; van Zomeren et al., 2004; 2008). Research has demonstrated that *group-based anger* (and people's subsequent engagement in action) can arise to the extent that people identify with a particular social group (e.g., they identify as a member of a disadvantaged group; van Zomeren et al., 2004; 2008). Related work proposes that this relationship can also work the other way around (Thomas & McGarty, 2009). That is, expressions of anger can contribute to people's perceptions that their grievances are shared, which, in turn, facilitates their identification with a group grounded in support for addressing the injustice (Thomas & McGarty, 2009; Thomas et al., 2012). This work provides an account for how social movements can *emerge* out of people's interactions about perceived injustice, rather than out of any pre-existing membership in a given group (i.e., groups categorized by race or gender; see also Bliuc et al., 2007; McGarty et al., 2009; 2014).

Taken together, this research suggests that Twitter users' expressions of anger serve

both an *identity-expressive* function (through communicating group norms that promote justice and equality), as well as working to further consolidate (or facilitate the development of) identities related to taking action to address gender equality (see Alberici & Milesi, 2016; Thomas et al. 2012; Klein et al., 2007; Turner-Zwinkels & van Zomeren, 2021). Further, identity expression (via anger about perceived injustices) can signal to others to join a given cause (Simon & Klandermans, 2001).

Expressions of solidarity

Alongside expressions of anger, the results provide evidence of users expressing solidarity with survivors of sexual violence, fellow activists, and leaders of the March 4 Justice cause. Topic 8 centres discussions regarding prominent advocates for survivors of sexual assault, Grace Tame and Brittany Higgins, who were both instrumental in sparking national discussions about the issue of sexual violence within federal politics and the broader country (Hill, 2021). Looking more closely at the top tweets in Topic 8 show Twitter users expressing admiration of – and solidarity with – both Tame and Higgins, who addressed protesters at the March 4 Justice rallies in Hobart and Canberra.

Research has shown that leaders play a key role in mobilising people to engage in collective action for social change (see Haslam et al., 2015; 2020; Subašić et al., 2011; 2014; Reicher et al., 2007). Leaders are considered *prototypical* members of the ingroup – that is, they are seen to exemplify the values, norms, and beliefs of a given group (Steffens et al., 2021). In this way, leaders are critical sources of social influence, as they are able to communicate to followers both *what* the group stands for, as well as *how* the group should act (Reicher et al., 2005; Reicher & Hopkins, 1996; Selvanathan & Jetten, 2020). Subašić et al. (2022, in press) suggests that it is the extent to which leaders are able to effectively express a sense of *who* the group is (that is, that leaders' messages clearly embody the goals and values of the group), that their influence in mobilising people for change can be realized.

Both Grace Tame and Brittany Higgins' speeches at the March 4 Justice protests emphasized a need for care, compassion, and solidarity with survivors of sexual violence. Higgins spoke to the crowd at the March 4 Justice event outside of Parliament House in Canberra, Australia: "Speak up, share your truth and know that you have a generation of women ready, willing and able to support you". Tame addressed protesters in Hobart, Australia, calling for an end to a culture of silence around sexual violence that allows it to go unchecked: "Evil thrives in silence. The start of the solution is quite simple – start making noise". In this way, both Tame and Higgins can be seen as leaders of the March 4 Justice movement, who are playing a central role in solidifying (or facilitating) people's support. In emphasizing the movement's central demands (an amplification of the voices of survivors and a need to stand in solidarity alongside them to demand change), Tame and Higgins encourage supporters of the movement to do the same. Twitter users expressing their solidarity with Tame and Higgins (both survivors of sexual abuse) can therefore be seen to be enacting the core message of the March 4 Justice movement.

Tweets in Topic 13 also capture users expressing solidarity with survivors of sexual violence by disclosing their own personal experiences of – or personal connection to – sexual harassment and assault. These tweets provide evidence for the notion that social media is an important source of support for victims of sexual violence (Schneider et al., 2019).

Demands for social change

Topics 3, 4 and 7 reflect people's demands for social and political changes to address the issue of gendered violence. Topic 3 is focused more broadly on the issue of sexism and misogyny within Australian federal parliament and the wider country. Top texts in Topic 3 include users describing a need for an independent inquiry into allegations of sexual abuse (one of the key demands outlined by the March 4 Justice organisers; Lee, 2021). Topic 4 is focused more specifically on users expressing concerns about the South Australian and New

South Wales police investigation into the allegations of historical rape made against former Australian Attorney-General Christian Porter.

Top terms associated with Topic 7 also show that this topic reflects people's demands for social change. Similar to other topics, top tweets in Topic 7 include users criticising the government's response to the issue of gender inequality. The top FREX term "cow" in this topic suggests that tweets included discussion and criticism of Liberal party senator Linda Reynolds, who was reported to have called former Liberal staffer Brittany Higgins a "lying cow" following Higgins' allegation that she was raped in Reynolds' office by a male staffer in 2019 (Hitch, 2021).

The instrumental use of Twitter

While a considerable number of topics in the corpus capture the identity-expressive functions of social media (via users communicating moral outrage and solidarity), there is also evidence that Twitter was used instrumentally by supporters to plan and organise the offline March 4 Justice protests. Alongside discussions related to the Australian legal system and Australian state politics, top tweets in Topic 12 provide evidence of Twitter users organising commuting to and from the protests in capital cities and towns across Australia, as well as spreading information about the time and date of the rallies. These findings are consistent with research by Kende et al. (2016), who propose that social media can serve both a "social affirmation" purpose (e.g., through the consolidation and expression of group identities), and an "instrumental" purpose, whereby the affordances of social networking platforms are harnessed to disseminate information and organise action. Organisers of the March 4 Justice protest used the hashtags "#March4Justice" and "#EnoughisEnough" to raise awareness of the campaign and convey their key demands. Lüders et al. (2022) notes that hashtags serve both an identity-expressive function as well as an instrumental one by

allowing users to identify their own position on a particular issue, as well as connect with like-minded others to plan and coordinate action (see also Barron & Bollen, 2021).

Discussions regarding the inclusivity of #MeToo

Topic 11 provides evidence of disagreements among supporters of the March 4 Justice movement, highlighting how advocates for a given social or political issue are often not a homogenous group. Rather, broad political movements (such as feminism) involve supporters from a number of different subgroups (e.g., liberal feminists versus socialist feminists), each with different ideologies and ideas about how best to achieve social change (Kerner, 2017; see also Thomas et al., 2019). Tweets in Topic 11 show users expressing concerns about the inclusivity of the #MeToo movement – in particular, the need for an intersectional movement that considers the disproportionate rates at which trans women and women of colour experience sexual violence. This particular finding aligns with previous work that notes that "#MeToo" tends to centre the voices of cisgender, white, heterosexual, able-bodied women. As a result, the movement has been criticised for excluding women who are most vulnerable to sexual violence (i.e., trans women, women of color, women with disabilities; Trott, 2021; Williams, 2021). In line with these discussions, Topic 13 provides evidence of users expressing concerns about the disproportionate rates of sexual violence experienced by Aboriginal women and children.

What is the nature of online opposition to March 4 Justice?

Defensive strategies designed to downplay the issue of men's violence against women

The results of the topic model provide evidence that expressions of support for #March4Justice co-existed alongside defensive backlash that sought to downplay (or outright deny) the issue of men's violence against women. Much of this backlash was captured in Topic 10, which included users discussing the impact of sexual assault allegations. Closer examination of top tweets in this topic reveal that opponents commonly questioned the

legitimacy of victims' claims and mobilised concerns about the impact of false rape allegations on the assumed perpetrators. In response, supporters stressed the rarity of false rape allegations and the adverse impacts of sexual violence on victims. The fact that Topic 10 was the most prevalent topic in the corpus (8.1%) provides support for the notion that disclosures of sexual assault are often highly contested (Boyle & Rathnayake 2019; Fileborn & Philips, 2019; Weiser et al., 2017).

Myths regarding false rape allegations are well documented in research across the disciplines of law, feminism, and psychology (Belknap, 2010; Burt, 1980). Research has demonstrated that survivors of sexual assault are often accused of falsifying events, despite empirical evidence documenting that false allegations are rare and account for between only 2-10% percent of reported assaults (Lisak et al., 2010). Evidence of concerns about false allegations in the March 4 Justice corpus are consistent with research suggesting that opponents of gender equality often push back against the movement to address gendered violence by claiming that men are being unfairly victimised by "vindicative women" (Flood et al., 2021; Gruber, 2009, p. 598).

The social psychological literature suggests that men's claims to victimhood can arise due to concerns about their group's moral image (Kahalon et al., 2019; Sullivan et al., 2012). In order to restore their group's moral reputation, men can engage in *competitive victimhood* by claiming it is men (rather than women) who are victimised (Blais & Dupuis-Déri, 2012; Boyle & Rathnayake, 2019; Flood et al., 2021, p. 396; Nicholas & Agius, 2018, p. 44; Noor et al., 2012; Sullivan et al., 2012). Thus, evidence of competitive victimhood in the corpus suggests that users may have been attempting to restore a threatened moral image caused by discussions about men's role as the perpetrators of sexual violence (see Sullivan et al., 2012).

Topic 11 provides further evidence that opponents of the March 4 Justice movement were concerned about the impact of sexual assault allegations on men's moral and social

reputation. Key terms for Topic 11 – "women", "white", "male" – suggest a focus on discussions about gender and racial identity. Indeed, top tweets in this topic reveal that users are employing a range of defensive rhetoric designed to downplay the issue of men's violence against women. For example, one user writes: "Gee ..Great time to be a White Male. Did or do all these women really hate their fathers and husbands that much? Oh. it is not all males then ? Just a few bad apples ? 99 % of men do not rape or bash women, please stop tarring all men with the same brush. Adjust your demands". This tweet reflects another defensive strategy that can be used to deflect against moral image threats: suggesting that only a few deviant ("bad apple") members of the ingroup are responsible for harmdoing (see Bilali & Vollhardt, 2019; Marques et al., 2001). By blaming a small minority of the ingroup, men are able to deny how gender inequality is culturally embedded and reproduced within social structures and institutions. That is, by claiming that they do not *personally* perpetrate sexual harassment or assault, men position the issue of sexual violence as an individual problem rather than a collective one – which works to derail important conversations about the structural nature of violence against women (Flood, 2018; Lanius, 2019). The tweet included above also suggests that women who support the #MeToo and #March4Justice movement are misandrists (or "man-haters"; Ging, 2017) – in line with research noting that Men's Rights Activists frequently position any form of feminist movement as prejudiced against men (Ringrose & Lawrence, 2018).

General Discussion

The current chapter sought to provide an overview of the nature of online support and resistance to the movement to address gender-based violence within Australia. Overall, the findings suggest that online interactions regarding the March 4 Justice movement can be understood as an intergroup conflict between supporters and opponents of the movement to address gender-based violence. Supporter identity was enacted via users' expressions of

anger at the perceived inaction of the government and solidarity with survivors of sexual violence. In contrast, opponent identity was expressed via anti-feminist rhetoric and defensive strategies that sought to minimize or deny the issue of men's violence against women.

Online Support for March 4 Justice

The results of the topic model suggest that Twitter was an important platform for supporters of gender equality to express their moral outrage at the actions of Australian political leaders and institutions, communicate solidarity with survivors of sexual violence, and demand social change. Further, supporters used Twitter as a tool to raise awareness of the March 4 Justice rallies across Australian towns and cities. There was also evidence that Twitter was a site of contestation between supporters – who discussed issues related to the inclusivity of "#MeToo". Users questioned which voices are privileged in discussions about violence against women, and which voices are at risk of being excluded.

These findings are in line with previous research demonstrating that social media is an important platform for people to mobilise for social change through airing grievances, negotiating potential forms of action, and building and expressing solidarity with fellow supporters and leaders of the cause (Gleeson & Turner, 2018; Mendes & Ringrose, 2018; Smith et al., 2018; Thomas et al., 2022a). Online interactions with like-minded others allow people to come to see that their grievances are *shared*, and that, together, they can do something to change the current state of affairs (Thomas & McGarty, 2009; Thomas et al., 2012; 2014; 2019).

Online Opposition to March 4 Justice

The results of the topic model suggest that opposition to March 4 Justice was largely characterized by attempts to defend men's moral reputation. Tweets included a range of defensive strategies – victim-blaming, minimizing the structural nature of gender-based

violence (by claiming that sexual violence is only perpetrated by a deviant few), and competing with women for victim status (by positioning men as the victims of false rape allegations).

Together, the findings suggest that concerns about men's moral image may play a key role in motivating people's engagement in reactionary action designed to resist the movement for gender equality. To date, the literature on reactionary forms of collective action has predominantly focused on how advantaged group members' concerns about their group's privileged position in the status hierarchy can lead them to resist any real (or perceived) challenges to the status quo (Becker, 2020; Choma et al., 2020; Jost et al., 2017; Reicher & Ulusahin, 2020; Stanley et al., 2019). However, the current findings suggest that, in the context of discussions about sexual violence against women, men may feel threatened by the idea that their group's moral reputation is under attack, and this, in turn, may motivate their counteraction to convince others of their group's morality (Sullivan et al., 2012; Young & Sullivan, 2016).

Research has demonstrated that advantaged group members' need for positive moral image (and subsequent defensiveness) can reduce their support for policies or actions designed to secure rights for the disadvantaged group (Kahalon et al., 2019; Kende et al., 2020; Teixeira et al., 2020). However, to date, research has not explicitly tested whether morality concerns and defensiveness may motivate advantaged group members' *countermobilization* in the form of action to promote the rights of their own group. Thus, it would be important for future research to explore how men's need to defend their group's morality shapes their support for progressive (versus reactionary) forms of social change.

However, it is important to acknowledge that the current study did not include measures of Twitter user's gender identity. As a result, the present findings are unable to speak to the relationship between aspects of Twitter users' social identities (e.g., advantaged

or disadvantaged group membership) and their support (versus opposition) to gender equality. While research has predominantly focused on men's resistance to feminist efforts, women too oppose progress towards gender equality (e.g., women who viewed #MeToo as a "witch hunt" against innocent men; Mikołajczak et al., 2022, p. 3). Therefore, future research could use machine learning techniques to capture users' gender identity, in order to examine the influence of gender on expressions of support (versus opposition) to the movement for gender equality online (Pathak et al., 2021). Further, recent research has highlighted the relationship between conservative ideologies and engagement in reactionary forms of collective action (Becker, 2019; Choma et al., 2020; Lisnek et al., 2021). Accordingly, future studies could also seek to capture Twitter users' political orientation and explore if a similar pattern of results is found in naturally occurring online debates about gender equality. Structural topic modelling would lend itself particularly well to this approach as it is designed to include "document-level characteristics" (that is, continuous or categorical variables in a given data set; Roberts et al., 2014).

The current study also faces limitations associated more generally with taking a topic modeling approach – namely, that interpretation of topics reflects researcher's subjective judgment (Jiang et al., 2021). Further, as a form of unsupervised machine learning, topic models tend to only capture the most identifiable aspects of a given corpus and can miss identifying themes that may be of theoretical importance (Jagarlamudi et al., 2012; Watanbe & Zhou, 2020). Future research should therefore seek to complement a topic modeling approach with a more in-depth qualitative analysis of tweets than the one provided in the present chapter.

Conclusion

The current study provides novel insights into how support and resistance to the movement for gender equality within Australia was expressed online. The findings suggest

that online interactions surrounding March 4 Justice involved a conflict between users

regarding the impacts of gender-based violence. While supporters emphasized a need to stand

in solidarity with survivors of sexual violence and hold powerful institutions and leaders to

account, opponents sought to deflect from the issue by positioning men as the victims of false

allegations. Given the prevalence of attempts to defend men's moral character in tweets from

opponents, it would be important for future research to explicitly test whether men's

intentions to support or oppose feminist efforts are shaped by their need to protect their

group's moral image.

Chapter 3

#MeToo, #MenToo: How Women's Accusations Shape Men's (Counter)Mobilization

Abstract

Online disclosures of victimisation from women (e.g., via movements such as #MeToo) are often met with backlash from men (e.g., #MenToo, #NotAllMen). Drawing on the needs-based model of reconciliation, the current research examined whether counter-claims from men are influenced by the perception that their group's moral image is under threat (and their subsequent need to defend against this threat). Three pre-registered studies ($N = 733$) found that men's social identification was associated with reduced intentions to act for women's rights and was positively related to their intentions to advance the rights of their own group – effects mediated by men's need for morality and defensiveness. Three manipulations that sought to vary levels of threat did not affect this process. Overall, the findings suggest that defensiveness may not only present a barrier to men engaging in solidarity-based action for gender equality but can also increase their intentions to participate in counter-movements designed to promote the rights of their own (privileged) group.

#MeToo, #MenToo: How Women's Accusations Shape Men's (Counter)Mobilization

Social media provides an important platform for women to raise awareness of gender inequality and mobilise for social change. The #MeToo movement gave women the opportunity to share their experiences of sexual harassment and assault and highlighted the pervasive nature of sexual violence against women (Kunst et al., 2019). Although #MeToo has garnered support from some men, (e.g., #HowIWillChange; PettyJohn et al., 2019), it has generated backlash and resistance from others (Boyle & Rathnayake, 2020l; Flood, 2019; Flood et al., 2021; Lisnek et al., 2021). #MenToo and #NotAllMen are two common hashtags used to respond to women's disclosures to assert that men are also the victims of sexual violence, and that not all men are perpetrators of sexual harassment and assault (Boyle & Rathnayake, 2020).

The current research aims to contribute to the growing literature on advantaged group members' resistance to social change and their participation in reactionary action designed to preserve the status quo (Becker, 2020; Jost et al., 2017; Osborne et al., 2019; Selvanathan et al., 2020b; Wilkins et al., 2015). Specifically, I explore the link between men's social identification (as members of a structurally advantaged group responsible for perpetrating harm) and their intentions to support action for women's rights *or* action designed to promote the rights of their own group. I propose that men's social identification will be associated with concerns about their group's moral image in the context of sexual harassment allegations, which will subsequently predict their defensiveness. I test whether defensiveness, in turn, is associated with men's *reduced* support for actions designed to promote women's rights, and *increased* support for counter-action intended to advance the rights of their own group.

#MeToo, #MenToo: Online movements and counter-movements

The phrase "Me Too" was first used online by civil rights activist Tarana Burke in 2006 to raise awareness of sexual violence against women – particularly against Black women and girls (Bhattacharyya, 2018; Gómez & Gobin, 2020). The movement gained global recognition in 2017 when actress Alyssa Milano used the hashtag "#MeToo" to encourage others to come forward and share their own experiences of sexual harassment and assault. In the days following Milano's tweet, the hashtag was used over 12 million times (Jaffe et al., 2021).

#MenToo emerged in direct response to #MeToo to argue that men are also the victims of sexual harassment and sexual violence. Alongside #MenToo, #NotAllMen was used to assert that not all men are perpetrators of sexual harassment and violence against women. While supporters of #MenToo allege the movement simply represents an appeal to inclusivity, many argue that such reactions reflect men's attempts to derail discussions regarding violence against women (Flood 2019; Flood et al., 2021; PettyJohn et al., 2019; Nicholas & Agius, 2017; Okuyan & Vollhardt, 2022; Wilkins et al., 2017).

Throughout history, disadvantaged groups' efforts to challenge inequality have met backlash and resistance from members of the advantaged group (Flood, 2021; Faludi, 1991; Mansbridge & Shames, 2008). Research suggests backlash from dominant group members can arise due to their desire to maintain a position of power and privilege, leading them to resist any real or perceived challenges to the status quo (Sidanius & Pratto, 1999; Jost & Banaji, 1994; Jost et al., 2004). Indeed, recent research in social psychology has focused on political conservatism (Becker, 2020; Lisnek et al., 2021), right-wing authoritarianism and social dominance orientation (Choma et al., 2020; Okuyan & Vollhardt, 2022; Stanley et al., 2019), system justification (Kende et al., 2020; Osborne et al., 2019), and dominant group members' perceived loss of power (Reicher & Ulusahin, 2020), as important predictors of

reactionary action designed to protect the advantaged group members' status. Recent work by Rivera-Rodriguez et al. (2022) demonstrated how men's perception that masculinity was losing social currency was positively associated with threat regarding men's status, which, in turn, increased their opposition to feminist social movements.

However, men's responses to women's disclosures of sexual violence online are often characterized by concerns about their groups' moral image. That is, while perceived threats to the advantaged groups' status and power play an important role in influencing resistance to progressive social change, a desire to defend one's groups' moral reputation may also motivate participation in reactionary forms of action. Indeed, counter-protests from advantaged group members primarily emerge in direct response to information that their group is responsible for harmdoing (i.e., a threat to their need for positive moral identity; Branscombe et al., 1999; Shnabel & Nadler, 2008; Sullivan et al., 2012). Further, advantaged group members' counter claims are often characterized by attempts to exonerate their group of moral responsibility through the use of defensive strategies: by morally disengaging from the wrongdoing (e.g., denying their groups' role as perpetrators; Bilali, 2013; Bilali et al., 2012); minimizing the severity of the wrongdoing (Bilali et al., 2012; Leidner et al., 2010); blaming the victim(s) (Bandura, 1996; 1999), and/or establishing that their group has suffered *more* than the disadvantaged group (*competitive victimhood*; Noor et al. 2012; Sullivan et al., 2012; Young & Sullivan, 2016).

Accordingly, in the present research I propose that defensiveness – that is, the various strategies people employ to protect against threats to their personal or group identity – may shape men's support for social movements for or against social change. I seek to extend previous research on the relationship between moral image concerns and advantaged group members' support for the disadvantaged groups' action (see Kende et al., 2020; Teixeira et al., 2020) by also considering how morality concerns may motivate advantaged group

members (men) to engage in collective action on behalf of their *own* group. Thus, my approach examines men's intentions to engage in both progressive (*system-challenging*) and reactionary (*system-supporting*) forms of collective action, side by side (Becker, 2020; Jost et al., 2017; Osborne et al., 2019; Wilkins et al., 2015). Below, I draw on the insights of the social identity approach (Tajfel & Turner, 1979) and the needs-based model of reconciliation (Shnabel & Nadler, 2008) to outline how men's social identification drives their need for positive moral image (and thus, should motivate their need to defend their ingroup following women's accusations of harmdoing).

What drives defensiveness: Advantaged groups' need for morality

According to the needs-based model of reconciliation (Shnabel & Nadler, 2008), victims and perpetrators have conflicting psychological needs that are impaired following a conflict. Victims experience a threat to their power, resulting in a need for acknowledgement of the wrongdoing and affirmation of their worth. Conversely, perpetrators experience a threat to their moral-social identity – leading them to experience a heightened need for morality and forgiveness from victims. Previous research has applied the needs-based model to understand the psychological needs of groups characterized by structural inequality, based on the idea that "victims" and "perpetrators" correspond in many ways to disadvantaged and advantaged groups, respectively (Siem et al., 2013; Shnabel et al., 2013b; Kahalon et al., 2019; Aydin et al., 2019).

Importantly, it is not always the case that being a member of an advantaged group results in a heightened need for morality. Social Identity Theory (Tajfel & Turner, 1979) outlines how people derive a sense of meaning and positive distinctiveness from their group memberships, providing a strong motivation to be part of groups that are considered moral (Leach et al., 2007a; Paulhus & John, 1998; Wojciszke, 1994). In the context of groups characterized by structural inequality, it follows that advantaged group members' need for

morality is shaped by the extent to which they socially identify with that group (that is, they view their membership with the advantaged group as an important part of their self-concept; Branscombe et al., 1999). Strongly identified group members are most likely to experience a threat to their morality when they are reminded of their groups' role in past wrongdoing (e.g., Peetz et al., 2010). Indeed, Branscombe and Miron (2004) suggest perpetrator groups will primarily experience a threat to their morality when they encounter information that highlights their group's culpability for wrongdoing, and when there are no exonerating circumstances regarding who is responsible.

Online campaigns from disadvantaged groups commonly draw attention to the past (or ongoing) harms committed by members of the advantaged group and are thus likely to constitute a significant threat to their need for moral acceptance (Teixeira et al., 2020). Indeed, Kende et al. (2020, p. 2) demonstrated that both supporters and opponents of #MeToo perceived the movement as categorizing men as the perpetrators of sexual harassment. Further, recent research from Lisnek et al. (2021) found that perceived increases in women's "voice" (i.e., women bringing awareness to sexual assault) led both men and women to perceive greater victimisation of men.

Taken together, these findings suggest that men will experience a threat to their moral identity in response to online disclosures of sexual harassment from women to the extent that their group is implicated as responsible for women's ongoing victimisation. Accordingly, I propose the relationship between men's social identification and their need for morality will vary depending on the extent to which women's disclosures accuse men as the perpetrators of sexual harassment.

The Present Research

In three experiments I examine the relationship between men's social identification and their intentions to act as allies and engage in solidarity-based action for women – versus

their intentions to participate in counter-action to protect their own group (a form of system-supporting or reactionary activism; Becker, 2020; Osborne et al., 2019). I examine whether these two distinct forms of protest are both shaped by men's need for morality and defensiveness (see Figure 3 for my full conceptual model). Based on the key tenets of the needs-based model, I test whether men's need for morality arises as a function of their commitment to their group membership. I expect this association will occur *especially* when women explicitly accuse men as the perpetrators of wrongdoing (via an online accusation of sexual harassment; see Figure 3). Men's need for morality should be associated with a need to defend their group's morality (i.e., defensiveness).
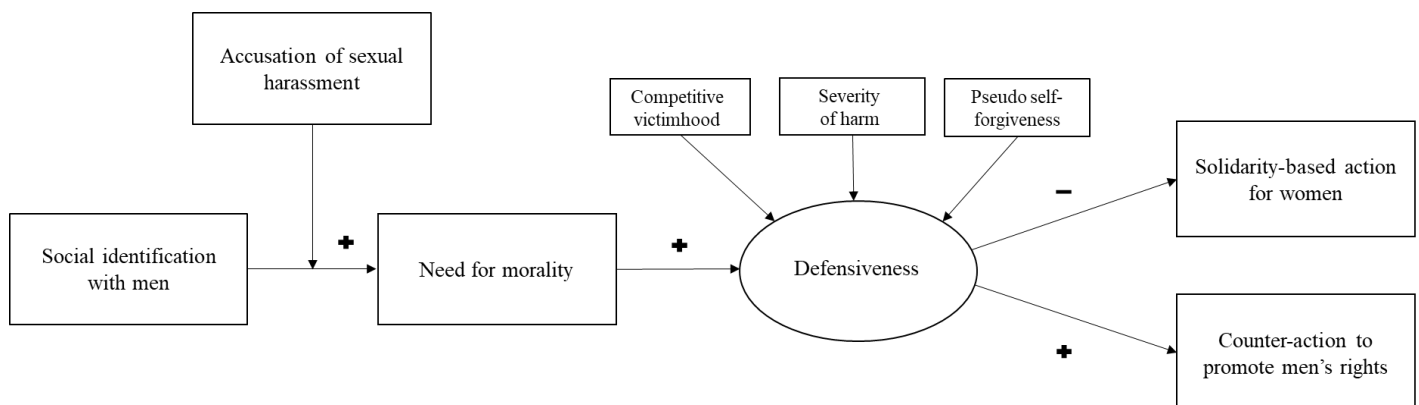
Figure 3. Conceptual model

Based on common ways groups can defend against threats to their identity identified in both the interpersonal and intergroup literatures, I operationalized men's defensiveness as the extent to which they competed with women for victim status (competitive victimhood; Noor et al., 2012; Sullivan et al., 2012), perceived men as experiencing *more* harm as a result of sexual harassment and its consequences compared to women (Bilali et al., 2012), and engaged in *pseudo self-forgiveness* (through minimizing harm, denying wrongdoing, and derogating the victim group; Fisher & Exline, 2006; Hall & Fincham, 2005; Woodyatt & Wenzel, 2013). Responses to each of these variables were parcelled together and modelled as reflective indicators of defensiveness. This approach allows for consideration of the shared or

common underlying construct (defensiveness), transcending individual literatures on (for example) pseudo-self-forgiveness and competitive victimhood per se. Statistically, the parcelling approach adopted here creates a more parsimonious model and also addresses measurement error (rather than using all items separately; Hall et al., 1999).

Finally, I test the association between men's defensiveness and their intentions to participate in action for women, or action to promote the rights of their own group (men). Recent research by Hässler et al. (2022) found that satisfying advantaged group members' need for acceptance was associated with their increased support for social change. Further, past research in the context of intractable violent conflicts showed group members' engagement in competitive victimhood led them to negatively evaluate the outgroup's protest (Demirdağ & Hasta, 2019). Accordingly, I expect that men's defensiveness will be associated with less willingness to engage in solidarity-based action for women's rights (see Figure 3). I also investigated the effect of defensiveness on men's intentions to engage in collective action to advance the rights of their own group. If the hypotheses are supported, it would show that the same process of defensiveness explains variation in commitment to actions that promote justice for women (negative effect) and men (positive effect).

The current studies used (fabricated) online content to explore men's responses to disclosures of sexual harassment from women. I tested my theoretical model (Figure 3) in all three studies using Multigroup Structural Equation Modelling (SEM). Multigroup SEM allows for testing complex mediation models across groups (Yuan & Chan, 2016). Study 2 examined whether an explicit accusation of sexual harassment moderated the relationship between men's social identification and their need for morality. Study 3 employed a different threat manipulation by presenting some men with information designed to affirm their group's morality following an accusation of sexual harassment. Study 4 manipulated the salience of men's social identification – as well as whether they engaged in a self-affirmation

intervention designed to buffer against identity threats – in order to experimentally test the influence of men's group membership and an accusation of harmdoing on their need for morality, defensiveness, and their intentions to participate in collective action (for women and men). All three studies were pre-registered on the Open Science Framework and I note any deviations from the preregistration documents[1].

Study 2

Participants were randomly allocated to read a bogus tweet that explicitly implicated men as the perpetrators of sexual harassment in the workplace (accusation condition) or not (no accusation condition). Social identification with men, need for moral acceptance, defensiveness, and collective action intentions were then assessed. I expected men's social identification would predict their need for morality when they were presented with an accusation of sexual harassment from women. Men's need for morality should be associated with defensiveness, which should, in turn, negatively predict men's willingness to engage in collective actions for women's rights. I also tested whether defensiveness was associated with men's intentions to engage in action to promote the rights of their own group (see Figure 3).

Method

Participants. Participants were 200 North American men ($M_{age} = 37.99$, $SD_{age} = 12.99$) recruited online via Amazon's Mechanical Turk for convenience (TurkPrime; Litman et al., 2017). Two participants were removed due to incomplete data, leaving 198 participants in the focal analyses. A sensitivity analysis in pwrSEM v 0.1.2 (Wang & Rhemtulla, 2021) for parameter estimation in structural equation modelling showed that the final sample of 198 participants was sufficient to detect an indirect effect ($f = .09$, i.e., a small effect) of social identification on collective action intentions via need for morality and defensiveness, assuming an alpha of .05 and power of 0.80. Participants were predominantly White (88%) and heterosexual (89%) and 55% were bachelor's degree educated or higher. Men reported

their political orientation just below the middle (*M* = 3.60, *SD* = 1.79) of a 7-point scale ranging from 1 = *Strongly Liberal* to 7 = *Strongly Conservative*.

Procedure. Participants completed a survey titled "Responses to Online Information" and were told the study was interested in how people respond to information they encounter on social media. To control for potential order effects, the measure of social identification was counterbalanced with half of the participants completing the measure first and the other half completing the measure at the end of the study. The order in which participants completed the measure of ingroup identification did not impact (directly or in combination) on their levels of identification, or on the other main variables of interest.

Participants were randomly assigned to one of two conditions (accusation/no accusation) where they were presented with a tweet. They were told the tweet had been randomly selected from an online database of tweets on a range of different social, political, and lifestyle topics. Following the manipulation, participants completed measures of morality-related needs, defensiveness, and collective action intentions to support women or men. Participants also completed additional exploratory measures unrelated to the focus of the current paper. For simplicity, I do not report these measures here, see the online supplementary file for details and results and the pre-registration can be viewed here: https://osf.io/fnxaq?view_only=6940e019f0114f04a32be8c56aee63dd. Finally, participants were debriefed and reimbursed USD2.00.

Accusation manipulation. To manipulate an accusation of men's wrongdoing, I varied the extent to which a tweet (from a fictional woman, Amanda) implicated men as the perpetrators of sexual harassment towards women in the workplace. In the accusation condition, the tweet accused men of being responsible for perpetrating sexual assault in the workplace and highlighted their need to address the problem: "I have experienced sexual harassment at work from men. It is a problem that needs to be addressed by men so that our

workplaces are safe and we can all get on with our jobs". In the no accusation condition, the

tweet did not explicitly blame men for perpetrating sexual harassment and framed sexual

harassment as a problem society must collectively address: "I have experienced sexual

harassment at work. It is a problem that needs to be addressed by society so that our

workplaces are safe and we can all get on with our jobs" (see Figure 4 for a screenshot of the

tweets across conditions).

Figure 4. Screenshots of the tweets used in the accusation condition (left) and the no

accusation condition (right).

Measures

Unless otherwise indicated, items were answered on a 7-point scale (1 = *strongly*

disagree, 7 = *strongly agree*).

Social identification. One item from each of Leach et al.'s (2008) five subscales of

group identification was used to assess men's social identification. Example items included

"The fact I am a man is an important part of my identity" and "I am glad to be a man", α =

.87.

Need for morality. Adapted from Shnabel & Nadler (2008), four items assessed

participants need for morality – their desire for acceptance and understanding from the

outgroup: "I wish that women would perceive men as moral", "I would like women to know

that men try to act fairly, and "I would like women to understand that men are not harsh

people", α = .88.

Defensiveness. Defensiveness was operationalised with a latent combination of

variables that, together, conceptually denote ways group members can defend their group's

moral image, either by downplaying their group's role in perpetrating harm, or through

attempting to gain back morality by claiming *their* group is the real victim.

Competitive victimhood was captured by the extent to which men believed their

group suffers greater injustice than women. Four items were adapted from Kahalon et al.

(2019), e.g., "Economically, men in America are discriminated against more than women",

"Men in America are now suffering more emotional pain than women", α = .92.

One item (adapted from Bilali et al., 2012) measured participant's perceptions of the

severity of harm inflicted on their own group (men) compared to women due to sexual

harassment in the workplace, "Which group experiences *more* harm as a result of sexual

harassment against women in the workplace?". Responses were measured on a bipolar scale

where 1 = *women* and 7 = *men,* such that higher scores reflected the perception that men

suffer *more* harm due to sexual harassment, compared to women.

Six items assessed the extent to which participants denied their group's role in the

wrongdoing and engaged in victim blaming (adapted from Woodyatt & Wenzel, 2013).

Example items include: "I think the person in the tweet was really to blame for what

happened", "I'm not really sure whether what men did was wrong", and "Men aren't the only

ones to blame for what happened", α = .82.

Collective action intentions for men and women. Participants indicated their

agreement with a series of statements involving intentions to act on behalf of their own group

(men) and women (solidarity-based action with women; Saab et al., 2015), e.g. "I intend to

advocate for equality for [women/men] in my own place of work", "I intend to raise

awareness of the issues that some [women/men] experience in the workplace by posting on social media", (men: $\alpha = .77$, women: $\alpha = .77$).

Results and Discussion

Scale inter-correlations and descriptive statistics across conditions for the key dependent variables are shown in Table 2.

Preliminary analyses. I first ran a regression using Hayes' PROCESS macro model 1 (Hayes, 2017) to test whether the accusation manipulation (0 = no accusation, 1 = accusation) moderated the relationship between men's social identification and their need for morality. There was no conditional effect of the manipulation on need for morality, based on levels of social identification ($B = -.16$, $p = .27$, 95% CI [-.45, .14]), nor was there a direct effect of the manipulation on participants' morality needs ($B = .74$, $p = .33$, 95% CI [-.75, .2.22]). Table 2 shows the means for need for morality were above the mid-point in both conditions, suggesting that, irrespective of the tweet they received, men reported a relatively high need to have their morality affirmed.

Multigroup Structural Equation Modelling. I used IBM SPSS Amos 25 to test my hypothesized moderated mediation model – that social identification will lead to a threat to morality (particularly when men are presented with an accusation of sexual harassment). Need for morality was expected to be positively associated with defensiveness, which should, in turn, be associated with men's reduced intentions to participate in collective action in solidarity with women. Further, I explored whether men's defensiveness was related to their intentions to act to promote their own group (men's rights).

Table 2. Summary of means, standard deviations, and 95% confidence intervals (across conditions), and intercorrelations for key variables (Study 2).

Variable	No accusation	Accusation	1	2	3	4	5	6	7
1. Social identification with men	4.83 (1.12) [4.60, 5.05]	4.98 (1.12) [4.75, 5.20]	-						
2. Need for morality	4.80 (1.32) [4.54, 5.07]	4.83 (1.23) [4.60, 5.04]	.40**	-					
3. Competitive victimhood	2.71 (1.50) [2.41, 2.99]	2.77 (1.58) [2.46, 3.11]	.22**	.36**	-				
4. Pseudo self-forgiveness	2.68 (1.14) [2.46, 9.11]	2.71 (1.10) [2.51, 2.95]	.25**	.34**	.68**	-			
5. Severity of harm	2.06 (1.68) [1.75, 2.40]	2.37 (1.60) [2.07, 2.70]	.21**	.18*	.44**	.59**	-		
6. Collective action intentions (for men)	3.55 (1.57) [3.23, 3.85]	3.29 (1.59) [2.98, 3.62]	.20**	.10	.11	.09	.16*	-	
7. Collective action intentions (for women)	4.14 (1.60) [3.79, 4.42]	3.66 (1.69) [3.32, 3.99]	-.06	-.17*	-.30**	-.40**	-.12	.63**	-

*Note. *p < .05. **p < .001

I employed the accusation manipulation (accusation/no accusation) as the grouping variable in a Multi-Group Structural Equation Model (SEM). Doing so allowed me to consider whether the specific path between social identification and need for morality was the same (or different) across the two groups of the accusation manipulation, therefore allowing for a test of the full moderated mediation model (Byrne, 2013). To assess model fit I report several widely accepted fit indices: the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Cut-off points for these fit indices were: .95 or higher for CFI; .08 or lower for SRMR, and values of .01, .05, and .08 indicating excellent, good, and acceptable fit, respectively for RMSEA (Hu & Bentler, 1999; Bentler, 2007). Indirect effects were computed using the indirect effects command in AMOS with 10,000 bootstrap samples (95% confidence intervals). I concluded the indirect effect was significant when the CI did not include zero. The constrained and unconstrained models were compared using the chi-square test statistic, appropriate for testing nested models. A non-significant chi-square value supports retention of the more parsimonious model (i.e., with the constrained paths).

I first fixed all paths to be the same across conditions to test the structural relationships between variables (see Model 1, Table 3). I then compared this constrained model (Model 1) with my hypothesized model (Model 2), in which the key path from social identification to need for morality was released (free to vary) across condition. Support for the moderating role of an accusation of wrongdoing would be obtained if the released model fit better than the model in which the path was constrained to be the same. All other paths were constrained to be stable across condition.

Table 3. Fit statistics of the two structural equation models (Study 2).

Model	Model description	χ^2	CFI	RMSEA	SRMR	Chi-square difference tests
Model 1	All paths are constrained	$\chi^2(32) = 51.15$	.96	.06	.06	-
Model 2	Hypothesized model: identification – need for morality path released	$\chi^2(31) = 49.99$	.96	.06	.05	$\chi_{diff}^2(1) = 1.16$

Table 3 shows the fit statistics for the two models tested. It can be seen that the difference between the models was not significant, $\Delta\chi^2(1) = 1.16$, $p = .28$, indicating the model did not fit better when the hypothesized paths were allowed to differ. Figure 5 shows that, consistent with my hypothesis, need for morality was positively associated with defensiveness, which, in turn, negatively correlated with intentions to participate in collective action on behalf of women. Unexpectedly, however, this process occurred irrespective of whether men were presented with an accusation of sexual harassment or not: in both cases social identification with men was positively related to their need for moral acceptance. Table 4 shows the indirect effects of identification on defensiveness and action to promote women (respectively), were significant, although the indirect effect on collective action to support men was not.
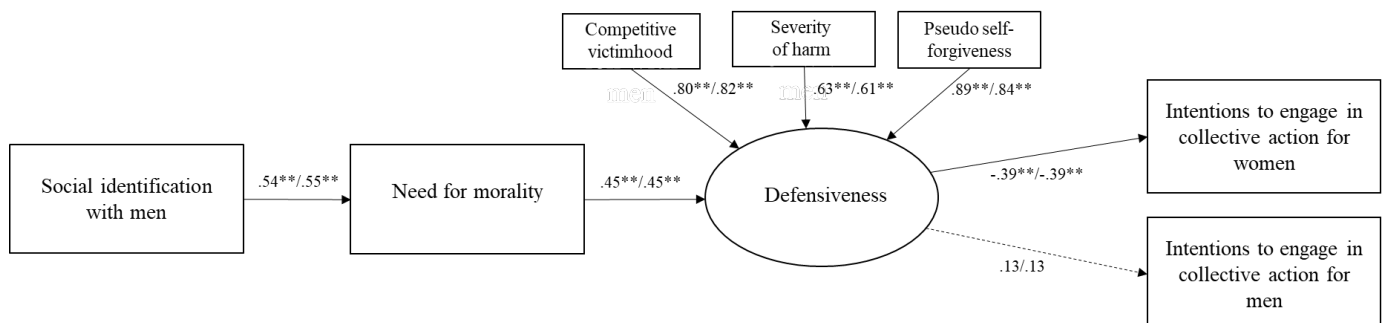

Figure 5. Standardized regression coefficients for the hypothesized structural model for Study 2. Values to the left of the slash represent the standardized weights in the accusation

condition and to the right of the slash represent standardized weights in the no accusation

condition. ** Denotes that the path is significant at $p <.001$. N.B. Residual error terms for

both collective action intention variables were allowed to covary.

Table 4. Indirect effects (Study 2).

Pathways	IE	SE	95% CI
Social identification → threat → defensiveness	0.16	0.05	0.07, 0.26
Social identification → threat → defensiveness → collective action intentions (for women)	-0.09	0.03	-0.16, -0.04
Social identification → threat → defensiveness → collective action intentions (for men)	0.03	0.02	-0.01, 0.07

Study 2 provides initial support for my conceptual model. Men's social identification

negatively predicted their intentions to advocate for women's rights via their need for their

group to be seen as good and moral and their engagement in defensive strategies (denying

men's role as the perpetrators of sexual harassment, competing with women for victim status,

and perceiving the consequences of sexual harassment to be harsher for men than women).

However, there was no evidence that the presence of an explicit claim of sexual harassment

moderated the relationship between social identification and men's need for morality. Results

indicated men's need for morality was high – above the midpoint – in both conditions.

Although the no accusation condition did not contain an *explicit* accusation of ingroup

wrongdoing, male participants may have felt that a tweet regarding sexual harassment (from

women) was implicitly holding their group accountable. That is, while the tweet in the

accusation condition assigned responsibility for both perpetrating and addressing sexual

harassment to men, the tweet in the no accusation condition did not explicitly assign

responsibility for the problem of sexual harassment. Men in this condition may have therefore

inferred blame for sexual harassment – and framing the responsibility of addressing sexual

harassment as a shared, societal one had no influence on mitigating this threat. Given this

possibility, in the next two studies I focused on manipulating a threat to men's need for positive moral identity by varying whether or not they read only about sexual harassment committed by men or whether this threat was mitigated. In Study 3, I sought to mitigate the threat to men's identity posed by an accusation of sexual harassment by providing some men with information about the moral actions of their group.

Study 3

In Study 3, all participants were presented with the same tweet used in the accusation condition in Study 2, however some were then also shown information designed to affirm their groups' morality. Moral affirmations have been successful in mitigating a threat to participants' collective identity (their collective self-esteem) in contexts of historical wrongdoing (Peetz et al., 2010). Further, recent research in the context of intergroup contact provides support for the use of moral affirmations in satisfying advantaged group members' need for morality and subsequently increasing their support for progressive social change (Hässler et al., 2022; see also Shnabel et al., 2013b).

Accordingly, in the current study I sought to create variation in men's need for morality by presenting some men with a fabricated news article that described men joining an online campaign to end sexual harassment. I expect that the relationship between men's social identification and their need for morality should be attenuated for men who are also presented with information affirming their group's morality alongside an accusation of ingroup wrongdoing. Thus, the mitigated threat condition should indirectly reduce men's defensiveness and, in turn, reduce their intentions to take action to promote justice for men and increase their support for collective action for women.

However, providing men with an article about their group mobilising on behalf of women does not just provide an example of their groups' morality, but also presents a descriptive norm of action that might separately influence their intentions to act in support of

women's rights. Thus, to control for these potential effects, the study design adopted two mitigated threat conditions. Both conditions used a fabricated news article (one described men taking online action to address sexual harassment, while the other described men as good parents and role models (this condition served as a control to disentangle the effects of moral affirmations and descriptive norms for action on the dependent variables; see Appendix A). This design also allows me to explore whether there are any differences in men's need for morality due to the *type* of affirmation strategy used. That is, whether the affirmation must be directly related to the threat posed to their group (i.e., by noting some men's efforts to fight against sexual harassment), or whether it can be unrelated (by affirming men's morality in another domain, i.e., as good fathers and role models; see Gunn & Wilson, 2011).

Method

Participants. 300 men (M_{age} = 37.16, SD_{age} = 11.85) from the USA were recruited via Amazon's TurkPrime crowdsourcing platform. Four participants were removed due to incomplete data, leaving a total of 296 participants for analysis. A sensitivity analysis conducted with pwrSEM v 0.1.2 showed that this final sample was sufficiently sensitive to detect an indirect effect of f = 0.13, assuming an alpha of .05 and power at .80. Again, participants were predominantly White (81%) and heterosexual (90.2%). Around 60% indicated that they had a bachelor's degree or higher. Mean political orientation sat just below the centre (M = 3.50, SD = 1.80) of a 7-point scale ranging from 1 = *Strongly liberal* to 7 – *Strongly conservative*.

Procedure. Study 3 was similar to Study 2 but adopted a three-cell between-groups design (accusation, mitigated threat, mitigated threat: control). Participants were randomly allocated to one of three experimental conditions. In the accusation condition, participants were presented solely with the accusatory tweet used in Study 2 (see Table 5). In the mitigated threat condition, participants were presented with the accusatory tweet and then

read about men joining a campaign to end sexual harassment against women (see Table 5 and

the supplementary materials for the articles used across conditions). To control for the

potential influence of descriptive norms on collective action, the mitigated threat control

condition presented men with the accusatory tweet alongside a news article that affirmed

men's status as good fathers. Following the manipulation, participants were then presented

with the same measures used in Study 2 (all measures formed reliable scales, α_{range} = .74 -

.92). For pre-registration document:

https://osf.io/4bmun/?view_only=8caa957371c24ac1b4ec1d4a58fea3e8.

Table 5. Study conditions (Study 3).

Condition	Stimuli	Stimuli excerpt
Accusation condition	Participants solely presented with accusatory tweet.	"I have experienced sexual harassment at work from men. It is a problem that needs to be addressed by men so that our workplaces are safe and we can all get on with our jobs"
Mitigated threat condition	Participants presented with accusatory tweet and news article: "Men join campaign to end sexual harassment"	*[screen shots of tweets embedded in article]* "I promise to speak out when I see sexual harassment taking place" "I will teach my sons to honor and respect women".
Mitigated threat control condition	Participants presented with accusatory tweet and news article: "Men leading by example: In praise of Fathers" article	*[screen shots of tweets embedded in article]* "My dad's provided for his family his whole life. He's the best male role model I know." "my father always let us know how much we were loved".

Results and Discussion

Preliminary analyses. Using Hayes' PROCESS macro model 1, I tested whether the effect of social identification on men's need for morality differed depending on experimental condition (0 = control, 1 = mitigated threat, 2 = accusation). There was no conditional effect of the manipulation on need for morality (B = .07, p = .34, , 95% CI [-.08, .22]) based on levels of social identification. Further, there was no direct effect of the mitigated threat manipulation on participant's need for morality (B = -.47, p = .23, , 95% CI [-1.24, 0.29]). Table 6 and 7 display the descriptive statistics for the key variables. It can be seen that the manipulation did not reliably affect any of the other mediating or outcome variables.

Table 6. Means (standard deviations) [95% confidence intervals] for key variables, by experimental condition (Study 3).

Variables	Accusation (n = 101)	Mitigated threat (n = 97)	Mitigated threat control (n = 98)
Identification with men	4.92 (1.06) [4.72, 5.14]	4.82 (1.09) [4.61, 5.05]	5.17 (1.11) [4.95, 5.37]
Need for morality	4.82 (1.20) [4.56, 5.07]	4.93 (1.23) [4.69, 5.18]	5.16 (1.36) [4.90, 5.41]
Competitive victimhood	2.96 (1.47) [2.67, 3.26]	2.71 (1.52) [2.43, 3.00]	2.93 (1.60) [2.64, 3.26]
Severity of harm	2.24 (1.53) [1.95, 2.56]	2.16 (1.58) [1.86, 2.51]	2.37 (1.74) [2.05, 2.70]
Pseudo self-forgiveness	2.71 (1.02) [2.51, 2.92]	2.77 (1.22) [2.53, 3.02]	2.76 (1.22) [2.51, 3.00]
Collective action intentions (for men)	3.31 (1.59) [2.99, 3.61]	3.75 (1.57) [3.42, 4.06]	3.17 (1.67) [2.84, 3.50]
Collective action intentions (for women)	3.76 (1.70) [3.41, 4.08]	4.19 (1.54) [3.89, 4.48]	3.63 (1.67) [3.28, 3.96]

Table 7. Correlations for key variables (Study 3).

Variable	1	2	3	4	5	6	7
1. Social identification with men	-						
2. Need for morality	.42**	-					
3. Competitive victimhood	.21**	.38**	-				
4. Pseudo self-forgiveness	.25**	.34**	.69**	-			
5. Severity of harm	.23**	.32*	.58**	.52**	-		
6. Collective action intentions (for men)	.22**	.14*	.11*	.14*	.20**	-	
7. Collective action intentions (for women)	-.03	-.10	-.31**	-.30**	-.12*	.61**	-

*Note. *p* <.05 **p* <.001.*

Multigroup SEM. As in Study 2, I tested for moderated mediation using multi-group structural equation modelling in IBM SPSS Amos 25. The accusation manipulation was the grouping variable (accusation, mitigated threat, and mitigated threat control).

Table 8 shows the fit statistics for the two models tested. I first fixed all paths to be the same across groups to test the structural relationships between variables (see Model 1, Table 8), before comparing this constrained model with my hypothesized model (Model 2, Table 8), in which the key path from social identification to need for morality was released (free to vary) across condition. Support for moderation would be obtained if the released model fit better than the model in which the path was constrained to be the same. The paths between need for morality and defensiveness, as well as between defensiveness and collective action intentions remained constrained because these relationships were predicted to be stable across conditions.

Table 8. Fit statistics of structural equation models (Study 3).

Model	Model description	χ^2	CFI	RMSEA	SRMR	Chi-square difference tests
Model 1	All paths are constrained	$\chi^2(51) =$ 68.72	.97	.04	.05	-
Model 2	Hypothesized model: identification – need for morality released	$\chi^2(49) =$ 67.97	.97	.04	.05	$\chi_{diff}^2(2) = 0.75$

Contrary to my moderation hypothesis, the difference between the models was not significant, $\Delta\chi^2(2) = 0.75$, $p = .69$, indicating the relationship between identification and need for morality was unaffected by the presence or absence of threat mitigating content. However, as in Study 2, I found evidence consistent with the predicted mediation between social identification and reduced collective action intentions on behalf of the outgroup via men's need for morality and defensiveness (see Figure 6). In Study 3, men's defensiveness was not only associated with a reduction in their intentions to participate in collective actions for women, but was also positively related to their intentions to participate in action on behalf of their own group. Results for the overall model indicated the indirect effect of both social identification on defensiveness – and on men's collective action intentions (for both women and men) – were significant (see Table 9).

Table 9. Indirect effects (Study 3).

Pathways	*IE*	*SE*	95% CI
Social identification → threat → defensiveness	0.30	0.04	0.22, 0.39
Social identification → threat → defensiveness → collective action intentions (for women)	-0.13	0.03	-0.19, -0.07
Social identification → threat → defensiveness → collective action intentions (for men)*	0.05	0.03	0.00, 0.11

Note. * = path is significant at $p \leq .05$
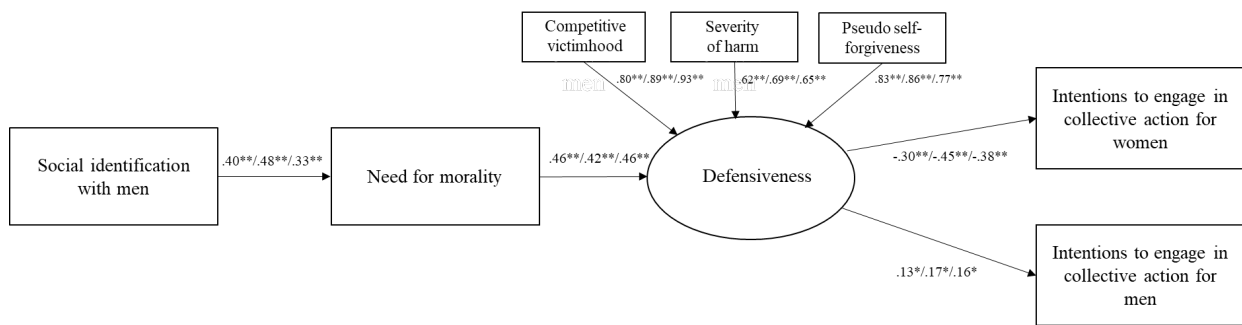
Figure 6. Standardized regression coefficients for the hypothesized structural model for study 3. Values to the left represent the standardized weights in the accusation condition. Values in the middle represent the standardized weights in the mitigated threat condition, and values to the right of the slash represent standardized weights in the mitigated threat condition and mitigated threat (control) condition. ** Denotes that the path is significant at $p < .001$. * d=Denotes that the path is significant at $p < .05$. N.B. The residual error terms for both collective action intention variables were allowed to covary.

Overall, the findings of Study 3 are consistent with those in Study 2: social identification as a man elicited men's need for morality, defensiveness, and was associated with reduced willingness to engage in collective action on behalf of women. However, again – contrary to predictions – these effects were produced under both conditions; when men were exposed to an accusation of ingroup wrongdoing, and also when they were presented with information designed to mitigate the effect of this accusation. In Study 3, men's defensiveness was also weakly positively associated with intentions to participate in action for men's rights, over and above the reductions in action in solidarity with women. This relationship did not attain significance in Study 2, although it is notable that the size of the effect was very similar (i.e., $\beta = .13/13$ in Study 1, $\beta = .13/.17/.16$ in Study 3).

As in Study 2, the scores for participant's need for morality were again high (above the midpoint) across all three conditions, suggesting that providing men with information about their efforts to address sexual harassment or their role as positive role models did not influence their need for moral acceptance. Given men's need for morality (and its flow on effects on defensiveness) ultimately stems from the degree to which their identity as man is self-defining (social identification), Study 4 explored the effect of experimentally manipulating men's social identity in combination with a threat to their morality via an accusation of harmdoing. In an attempt to create variation in men's need for morality across conditions, Study 4 utilised a self-affirmation task (McQueen & Klein, 2006) designed to protect against threats to one's personal or group identity.

Study 4

Study 4 tests experimentally the interaction between men's social identification and an accusation of wrongdoing. I adopted a self-affirmation intervention that has been found to buffer against the threat of an accusation of harm and reduce defensive responding, increase acknowledgement of past wrongdoing, and reduce outgroup derogation and prejudice in intergroup conflict settings (see Čehajić-Clancy et al., 2011; Fein & Spencer, 1997; SimanTov-Nachlieli et al., 2018). Self-affirmation theory (McQueen & Klein, 2006) proposes that when people spend time affirming positive aspects of their personal identity, this allows them to better cope with threats to their group identity (Sherman & Cohen, 2006; Steele, 1988). Accordingly, I expect self-affirmation should attenuate the relationship between men's social identification and their need for morality, as they have spent time bolstering their positive self-image. I expect this effect will be even more pronounced when participant's personal identity is salient, as they should be less attuned to information that may threaten their group's image when identification with their ingroup is low. Thus, a 2 (affirmation: no self-affirmation task, self-affirmation task) x 2 (identification: personal identity salient,

identification with men salient) between-groups design was employed. As in the previous two studies, men's need for morality, defensive strategies, and collective action intentions (for women and men) were measured variables.

Method

  Participants. The final sample comprised of 239 American men (M_{age} = 38.86, SD = 12.89, 93% heterosexual). I recruited participants via TurkPrime and paid them USD2.00 for their participation. One participant was excluded due to incomplete responding. A sensitivity analysis conducted using pwrSEM v 0.1.2 revealed that the final sample of 239 had sufficient power (alpha = .05, power = .80) to detect a small indirect effect (f = .03) of social identification on collective action intentions. 69% of participants indicated they were White, 18% indicated they were of Asian ethnicity, and 10% were Black. Around 63% indicated they held a bachelor's degree or higher. As in the previous two studies, mean political orientation fell just below the middle (M = 3.58, SD = 1.63) of a 7-point scale (1 = strongly liberal, 7 = strongly conservative).

  Procedure. As in Study 3, all participants were presented with the same fabricated tweet highlighting men's responsibility for maintaining and addressing sexual harassment in the workplace. However, in this study, participants were first randomly allocated to one of two experimental conditions to manipulate the salience of their personal or social identification (with other men). To manipulate the salience of men's social identity I adopted the procedure used by Haslam et al. (1999). In the social identification with men salient condition, participants were asked to identify and list three things that they and most other men (a) do often, (b) do rarely, (c) do well, and (d) do badly. In contrast, to make participants' personal identity salient, participants completed the same prompts but in relation to activities that they personally engage in. Following this, participants were randomly allocated to either receive a self-affirmation task (or not) adapted from previous research (see

McQueen & Klein, 2006). Participants were instructed to rank a list of values in order of personal importance, before being asked to write about 3-4 times their top-ranked value was of importance to them and made them feel good about themselves. Participants in the no affirmation condition did not complete any other task and were only presented with the accusatory tweet. The pre-registration document for Study 4 can be found here:

https://osf.io/h5wbn/?view_only=e8f087edc87345d2ac985cb1a39bcda6.

Measures

Participants completed the same measures used in the previous two studies (all measures formed reliable scales, α_{range} = .68 - .92). In this study, social identification was measured on a Likert scale from 1 (strongly disagree) to 9 (strongly agree) to capture more variability in responses. As in Studies 2-3, several additional exploratory variables were included; results are available in the online supplementary file.

<div align="center">

Results and Discussion

</div>

Preliminary analyses. An independent t-test was conducted in SPSS to test the effectiveness of the group salience manipulation. The manipulation was successful: men in the 'identification with men salient' condition reported greater social identification with men than participants in the 'personal identity salient' condition, $t(223.75)$ = -2.18, p = .03, d = 0.26 (see Table 10 for means).

Table 10. Means, standard deviations, [95% confidence intervals], and intercorrelations among measured variables (Study 4).

Variables	Accusation condition (N = 125)	Self-affirmation condition (N = 114)	Personal identity salient (N = 119)	Identity with men salient (N = 120)	1	2	3	4	5	6	7
1. Social identification with men	6.35 (1.39) [6.11, 6.58]	6.27 (1.58) [5.98, 6.56]	6.10 (1.64) [5.81, 6.37]	6.52 (1.64) [6.28, 6.75]	-						
2. Need for morality	4.93 (1.41) [4.67, 5.18]	4.70 (1.36) [4.46, 4.96]	4.64 (1.48)] [4.38, 4.92]	5.00 (1.27) [4.76, 5.21]	.47**	-					
3. Competitive victimhood	2.52 (1.58) [2.28, 2.80]	2.91 (1.55) [2.64, 3.21]	2.68 (1.56) [2.39, 2.93]	2.74 (1.52) [2.49, 3.02]	.34**	.39**	-				
4. Pseudo self-forgiveness	2.97 (0.99) [2.80, 3.14]	3.21 (1.00) [3.03, 3.39]	2.93 (0.97) [2.76, 3.12]	3.23 (1.00) [3.05, 3.41]	.25**	.34**	.68**	-			
5. Severity of harm	2.22 (1.59) [1.96, 2.51]	2.52 (1.66) [2.20, 2.81]	2.24 (1.65) [1.96, 2.54]	2.48 (1.60) [2.22, 2.77]	.09	.20**	.42**	.53**	-		
6. Collective action intentions (for men)	3.37 (1.78) [3.03, 3.68]	3.59 (1.60) [3.31, 3.89]	3.40 (1.68) [3.09, 3.68]	3.55 (1.71) [3.24, 3.86]	.27**	.32**	.15*	.16*	.01	-	
7. Collective action intentions (for women)	3.80 (1.75) [3.48, 4.10]	4.15 (1.61) [3.87, 4.45]	3.94 (1.73) [3.64, 4.28]	3.99 (1.66) [3.69, 4.31]	-.01	-.03	-.21**	-.22**	-.21**	.64**	-

Note. *p < .05. **p < .001. Social identification ranged from 1 (strongly disagree) to 9 (strongly agree).

Next, a two-way factorial ANOVA was used to test the effects of the group salience and self-affirmation manipulations on men's need for morality. There was a significant main effect of the group salience manipulation on men's need for morality $F(1, 235) = 4.29$, $p = .039$, $\eta^2 = .018$. However, there was no main effect of self-affirmation on morality needs, $F(1, 235) = 2.02$, $p = .156$, $\eta^2 = .008$ – or any interactive effects of group salience and self-affirmation on need for morality, $F(1, 235) = 0.55$, $p = .46$, $\eta^2 = .002$. Table 10 shows that self-affirmation had a direct effect ($p = .49$, $d = 0.26$) on men's engagement in competitive victimhood, however in the opposite direction to what was expected. Self-affirmation did not affect any of the other outcome variables, nor did it interact with the social identification manipulation to shape outcomes.

Multigroup SEM. As in Study 2 and 3, I tested my hypothesized conditional mediation model using Multigroup SEM. The self-affirmation manipulation was the grouping variable (self-affirmation condition, no self-affirmation).

Table 11. Fit statistics for the two structural models (Study 4).

Model	Model description	χ^2	CFI	RMSEA	SRMR	Chi-square difference tests
Model 1	All paths are constrained	$\chi^2 (32) = 61.04$	.94	.06	.05	-
Model 2	Hypothesized model: release identification – need for morality	$\chi^2 (31) = 60.48$	.94	.06	.05	$\chi_{diff}^2(1) = 0.56$

Table 11 shows the fit statistics for the two models. Model 1 shows the results for when all paths were held constant across groups. I compared this constrained model with the hypothesized model (Model 2), again allowing the path from social identification to need for morality to vary across conditions. As in Study 2 and 3, the difference between the structural models was not significant, $\Delta\chi^2(1) = 0.56$, $p = .45$, indicating that self-affirmation did not

moderate the relationship between social identification and need for morality. I therefore retained the more parsimonious model (Model 1) as the final model.

Figure 7 displays the standardized regression coefficients and Table 11 shows the model evidenced acceptable fit with the data. The salience of participant's identification (coded -1 for personal identity salient and 1 for identification with men salient) predicted need for morality, which in turn predicted defensiveness. Defensiveness was again modelled as a latent variable comprised of competitive victimhood, denial of harm and pseudo self-forgiveness, and was a direct predictor of men's intentions to engage in collective action for women and their intentions to engage in collective action on behalf of their own group. The paths between these variables remained constrained as these relationships were predicted to be stable whether participants were in the self-affirmation condition or not. Table 12 shows that, in line with Study 2 and 3, the indirect effect of social identification on defensiveness and men's intentions to engage in collective action for women's rights (respectively) was significant.
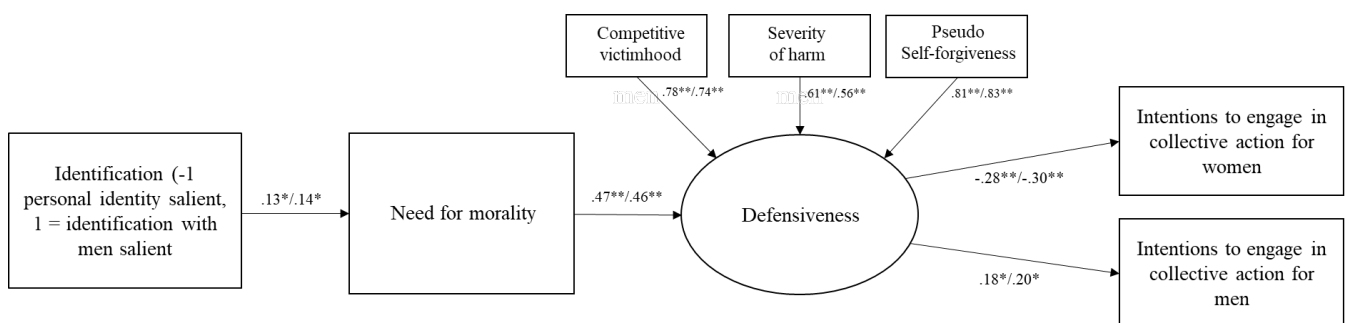
Figure 7. Standardized regression coefficients for the hypothesized structural model for Study 4. Values to the left of the slash represent the standardized weights in the no self-affirmation condition, and to the right of the slash represent standardized weights in the self-affirmation condition. ** Denotes that the path is significant at $p < .001$. * Denotes that the path is significant at $p < .05$. N.B. The residual error terms for both collective action intention variables were allowed to covary.

Table 12. Indirect effects (Study 4).

Pathways	IE	SE	95% CI
Social identification → threat → defensiveness	0.05	0.03	0.00, 0.10
Social identification → threat → defensiveness → collective action intentions (for women)	-0.03	0.02	-0.07, -0.00
Social identification → threat → defensiveness → collective action intentions (for men)*	0.02	0.01	-0.07, -0.00

Note. *Path is marginal at $p = .07$.

The results of Study 4 provide experimental support for the relationship between the salience of men's social identification and an increased need for moral acceptance. In line with Study 2 and 3, men's engagement in defensive strategies was not only negatively associated with intentions to act collectively for women's rights, but also positively predicted the likelihood they would act to promote the rights of their own group (S3-4).

General Discussion

Online interactions between advantaged and disadvantaged groups are pervasive in the 21st century, yet relatively little research has investigated how members of advantaged groups encounter, support, or resist online social change movements initiated by disadvantaged groups. Building on the insights of the social identity approach (Tajfel, 1982) and the needs-based model of reconciliation (Shnabel & Nadler, 2008), in the context of gender inequality between men and women I explored whether (a) men's social identification is associated with their reduced intentions to act collectively to support women and an increase in their intentions to advocate for men, (b) whether these effects are mediated by men's need for morality and defensiveness, and (c) whether this effect occurs when men are explicitly accused of sexual harassment from women.

Three studies provided evidence for my conceptual model (Figure 2), with some caveats. Social identification as a man was positively associated with men's desire for women

to accept their group as good and moral, which, in turn, was positively related to attempts to defend their group. Defensiveness was negatively associated with men's intentions to act collectively for women's rights (Studies 1-3), and positively predicted their intentions to act collectively for their own group (Studies 2-3). In Study 2 the relationship between defensiveness and men's intentions to act to support their own group did not reach significance. This may be because – compared to Studies 3 and 4 – Study 2 was not as well-powered to detect this effect. In Study 4, I manipulated the salience of men's social identification and found that, in line with the first two studies, men experienced a heightened need for morality (which was, in turn, positively related to defensiveness) when their identity with men was made salient, compared to men whose personal identity was made salient.

Unexpectedly, and across all three pre-registered studies, the relationship between men's social identification and need for morality was not influenced by an explicit accusation of ingroup wrongdoing. That is, neither manipulating the framing for who is responsible for committing and addressing sexual harassment against women (Study 2), nor attempting to mitigate this threat by affirming men's morality (Study 3) or participant's personal identity prior to this accusation (Study 4), altered the strength of the relationship between the importance of "being a man" (identification) and men's need for positive moral image.

Men's social identification reduced their intentions to act for women and increased their intentions to act for men via their need to defend their group

Results across all three studies found men's social identification led them to claim their group suffers more than women, perceive men as suffering *more* as a result of sexual harassment than women, and deny men's involvement and responsibility in perpetrating sexual harassment (pseudo self-forgiveness; Hall & Fincham, 2005) via a heightened need for morality. This pattern of results is consistent with Killian's (1985) concept of "stigma reversal" – the idea that highly identified advantaged group members can experience a threat

to their morality simply by virtue of their membership in a privileged group (see also Kahalon et al., 2019). This threat can lead them to engage in defensive behaviours in an attempt to convince others their group is good and trustworthy (Sullivan et al., 2012). The results of Study 4 provided further support for this notion. Instructing participants to focus on their group membership with men (rather than their unique personal identity) had a positive indirect relationship with men's defensiveness via an increased need to convince women of their morality.

Moreover, the findings demonstrated men's need to defend their group's morality negatively predicted their overall intentions to engage in online collective action for women's rights. This finding aligns with Kende et al.'s (2020) recent work examining support for the #MeToo movement, which found men's concerns about their moral reputation were associated with reduced campaign support. Further, research by Teixiera et al. (2020) found that advantaged group member's support for the disadvantaged group's protest decreased as a function of participants' concerns about their group's social image.

However, the present research goes beyond prior findings to explore the effects of morality concerns and defensiveness on men's mobilization to support disadvantaged group members, *as well as* their own advantaged group. Results showed men's defensiveness was positively associated with their intentions to participate in actions to promote the rights of their own group (S2-3), over and above the decreased commitment to support women (Kende et al., 2020; Teixeira et al., 2020). Thus, the results suggest that advantaged group members' need to protect their ingroup's reputation may not only act as a barrier to their participation in solidarity-based action with the disadvantaged group, but can also motivate them to act to advocate for their own group.

The findings compliment recent work highlighting the importance of considering collective action that aims to promote social change for disadvantaged groups *alongside*

collective action designed to preserve the existing social order (Becker, 2020; Jost et al., 2017l; Osborne et al., 2019; Selvanathan et al., 2020b). The findings contribute to social psychology's understanding of these actions by showing that defensiveness can contribute to both forms of action: on the one hand, reduce the likelihood that advantaged group members engage in solidarity-based action with the disadvantaged group (i.e., by supporting women's rights) while also increasing their intentions to engage in system-supporting action (i.e., by promoting men's rights). This paper is thus one of only a small number that considers both mobilization to support structurally advantaged *and* disadvantaged groups (see also Thomas et al., 2020). The results draw attention to the need to consider advantaged group member's need for morality and defensiveness as predictors of system-supporting action, therefore bringing together research on system-justification and collective action (Jost et al., 2017; Osborne et al., 2019) with research on group's psychological needs (Shnabel & Nadler, 2008; 2009; Nadler & Shnabel, 2015; Siem et al., 2013) and the defensive strategies groups can take to address those needs (see Bilali & Vollhardt, 2019, for a review). This synthesis seems particularly important given the rise of system-challenging action from advantaged groups, and how these actions often involve assertions of the ingroup's morality (e.g., claims to victimhood; Young & Sullivan, 2016).

The decision to conceptualize defensiveness as a latent factor further contributes to the literature by considering the similarities between the strategies groups can take to defend against threats to their identity. To date, the influence of such strategies on intergroup outcomes has predominantly been looked at independently from one another, e.g., as with research on competitive victimhood (see Young & Sullivan, 2016, for a review) or research examining perpetrator group's attributions of harm and responsibility (see Bilali et al., 2012). My analysis highlights the similarities between these processes. That is, they can arise due to advantaged group member's heightened need for morality, and together (as indicators of the

underlying latent factor of defensiveness) influence collective action intentions (on behalf of both the high-status and low-status group).

Women's accusations of victimisation did not moderate the relationship between men's social identification and need for morality

Across all three studies there was no evidence that the presence/absence of a direct or explicit accusation of wrongdoing (S1) – or attempts to mitigate this threat (S2 and 3) – influenced the relationship between men's social identification and their need for morality. These findings are inconsistent with previous research showing that providing group members with "mitigating" information (e.g., information that their group has made attempts at reparation, Peetz et al., 2010) reduces their need for morality. However, this research has largely been conducted in the context of violent intractable conflicts or historical memory for wrongdoing. For example, Peetz et al. (2010) manipulated German participant's collective self-esteem in the context of reminders of the Holocaust. Further, self-affirmation manipulations as a way of reducing defensiveness (and triggering prosocial outcomes) have been implemented in the context of relations between Israelis and Palestinians and Bosnians and Serbians (Čehajić-Clancy et al., 2011).

It may be that, in the context of gender inequality between men and women, men's need for moral acceptance is unaffected by information that attempts to mitigate this need. Given the frequency of interactions between men and women in personal, professional, and political life, it is possible that the link between men's ingroup identification and their need for morality is a chronically salient one (Ellemers & Barreto, 2003). As women become increasingly empowered to share their experiences of victimisation, the number of accusations implicating men as the perpetrators of harm has risen. As a result, attempts to manipulate an accusation of men's wrongdoing experimentally may have been overpowered

by the current socio-political context in which this research was conducted – where allegations of men's wrongdoing are widespread.

The results of the present research suggest that solutions to attenuating the indirect relationship between men's social identification and their need for defend their ingroup may lie in targeting men's social identification – that is, what it *means* to be a man. This could include attempts to align male identity with a "pro-gender equality" orientation (or *opinion*; Bliuc et al., 2007), e.g., via manipulating identity normative content to include men's expressions of support and engagement with women's rights. Further, endorsement of normative content by prominent majority group leaders has been shown to increase the likelihood group members adopt norms as central to their group identity (Haslam et al., 2015; see also Subašić et al., 2008). Future research should examine the influence of such identity (and leadership) manipulations on men's need for morality, and its flow on effects on defensiveness and men's collective action intentions.

Limitations and future directions

Adopting a latent measurement approach using multi-group SEM allowed me to test a complex set of hypotheses while taking measurement error into account. However, given the correlational nature of the data I cannot make strong claims regarding the direction of effects. Nevertheless, the ordering of variables in my analyses is consistent with past theory and research undertaken in the context of structural inequality and intractable conflict between groups (Noor et al., 2012; Shnabel et al., 2008, 2009, 2013b; Sullivan et al., 2012). Despite this, it is possible men's intentions to participate in collective action on behalf of their own group may motivate them to defend their group more strongly. Similarly, men's engagement in defensive strategies (e.g., denying that men have caused harm) may further heighten their need to convince women group of their morality – and this need may in turn reinforce men's identification with their ingroup.

An additional limitation of my SEM approach is that it does not allow for the consideration that different defensive strategies may be more or less likely to lead to men's support or resistance for women's protest. Future research should therefore consider the unique nature of different defensive strategies. For example, research could explore when and why advantaged group members will endorse one defensive strategy over the other, and the influence of each strategy on advantaged groups' support or opposition for action or policies to support the disadvantaged group.

Another limitation of the current studies concerns the measure of collective action intentions, which tapped into support for "men's rights" and "women's rights" broadly, without capturing the range of different reactionary and progressive actions men may engage in. It would be important for future work to explore, for example, how defensiveness may influence men's attempts to protect men from sexual assault allegations, or their support for programs designed to address 'toxic' masculinity (see Mikołajczak et al., 2022, p. 15).

Conclusion

Given the importance of action amongst advantaged group members (either 'allied' or 'in solidarity' with members of disadvantaged groups; Subašić et al., 2008, 2018; Louis et al., 2019), the current research aimed to contribute to understanding how members of structurally advantaged groups respond when they encounter accusations of victimisation from disadvantaged groups online. I demonstrate that defensiveness is not only a key barrier to men's engagement in solidarity-based action with women, but that it can also influence their intentions to advance the rights of their own group as a form of reactionary or system-supporting action. Importantly, this effect was driven by men's social identification and need for moral identity – and was not influenced by attempts to mitigate this need. The current paper argues for exploring the processes that underpin high-status groups' support *and* resistance to claims from low-status groups. Further research exploring both actions together

will allow for a more comprehensive understanding of the dynamics of online collective

action and its consequences.

Notes

[1] Deviations from pre-registration documents
i. Structural equation modelling (SEM) was used to test the hypothesized model rather than PROCESS as specified in the pre-registration document for Study 2.
ii. I did not initially pre-register defensiveness as a latent variable in Study 2 and 3. Defensiveness as a latent variable was pre-registered in Study 4.
iii. Collective action intentions (to promote [men/women]) were initially pre-registered as exploratory variables but are reported here as focal outcome variables.
iv. Study 2 and Study 3 reported accusation of harm as the predictor variable and social identification as moderator variable. For Study 4 and in the write up I have instead conceptualised social identification as the predictor variable and accusation of harm as the moderator.

Data availability
The data that support the findings of this study are available at
https://osf.io/bk4md/?view_only=b01e9e0f4a8446918f9fa639ff71b175

Chapter 4

Reducing Defensiveness and Increasing Action for Gender Equality

via Group Interaction

Despite numerous advancements in women's rights over the course of the last century, the issue of violence against women remains a global public health concern. Recent evidence indicates that around 1 in 3 women have experienced physical and/or sexual violence in their lifetime (World Health Organisation, 2021). Exposure to violence is even higher among black and ethnic minority women, sexual minority women, women with disabilities, and transgender women (Ali et al., 2022; Gómez et al., 2022; Ussher et al., 2022). In recent years, renewed protest from women (via movements such as #MeToo) have called on men to challenge sexist and misogynistic norms that give rise to violence against women. However, when confronted with women's allegations of sexual violence, many men have responded defensively (e.g., by claiming that the feminist movement victimises "innocent" men; de Maricourt & Burrell, 2022; Flood, 2019; Lisnek et al., 2022).

Defensiveness presents a significant barrier to men's engagement in the fight to address gender inequality. However, to date, little work has explored strategies aimed at mitigating defensive reactions as a potential means of engaging men in efforts to support women's rights. In the current chapter I draw on research from the social identity tradition (Tajfel & Turner, 1979) on the mobilising effects of group interaction on collective behaviour (McGarty et al., 2009; Postmes et al., 2005; Smith et al., 2015; Thomas & McGarty, 2009; Thomas et al., 2019). I propose that group interaction regarding strategies to promote justice and equality for women may provide a means of reducing men's defensiveness via the development of a social identity rooted in *acknowledging* – rather than denying – the issue of gendered violence. I explore whether engaging men in small group interactions about addressing sexual harassment and assault against women is associated with their increased

participation in solidarity-based action for gender equality. Specifically, I test whether the effects of group interaction on men's commitment to act is mediated by an increase in identification as a supporter of efforts to end sexual harassment and assault, and reduced defensiveness regarding the issue of gendered violence.

Defensiveness as a barrier to men's solidarity-based action for gender equality

Engaging men and boys in action for gender equality has been the focus of activists, academics, and policy-makers in recent years (Flood, 2019; Iyer & Ryan, 2009; Morgenroth & Ryan, 2018; Stewart, 2017; Ryan, 2022; Subašić et al., 2018; Sudkämper et al., 2020). Research suggests that male allies are important in supporting the movement for gender equality at both the micro (challenging norms and attitudes that promote sexism and misogyny) and macro (structural) levels of society (Drury & Kaiser, 2014; Selvanathan et al., 2020a; Williams et al., 2021). Further, research has shown that men who confront sexism are taken more seriously and viewed more positively than women who challenge gender-based discrimination (Drury & Kaiser, 2014).

Although men seem to play an important role in agitating for change to the status quo, many do not engage in efforts to address gender-based inequalities (Flood, 2019; Flood et al., 2020). Men's engagement in action for gender equality requires acknowledgement of their group's power, privilege, and history of oppression over women. This poses a unique challenge, as members of advantaged (or perpetrator) groups are strongly motivated to defend against information suggesting their group has done wrong (Bandura, 1996; Bilali & Vollhardt, 2019; Branscombe et al., 1999; Doosje et al., 1998; Knowles et al., 2014; Peetz et al., 2010). Indeed, defensive reactions from men to gender equality efforts are commonplace (Flood et al., 2020; Persson & Hostler, 2021). Men are more likely to deny the problem of gender inequality, blame women for the occurrence of sexual assault and harassment, and make claims of reverse discrimination in response to women's disclosures of violence (e.g.,

by arguing it is men who are now mistreated in society; Flood, 2019; Flood et al., 2020; de Maricourt & Burrell, 2021).

Research has shown that defensive reactions create a significant barrier to intergroup reconciliation efforts. The use of defensive strategies (such as denial, minimizing harmdoing, and competing with the outgroup for victim status) has been found to reduce intergroup forgiveness, trust, empathy, and support for collective action efforts on behalf of the disadvantaged group (Demirdağ & Hasta, 2019; Nadler & Shnabel, 2017; Young & Sullivan, 2011). Indeed, in Chapter 2 I provided evidence of defensive rhetoric being mobilized by Twitter users to oppose the #MeToo movement within the context of Australian politics. In Chapter 3, I demonstrated how, in the context of online sexual harassment claims from women, men's concern about their ingroups' morality positively predicted defensive construals regarding the issue of sexual harassment and assault against women. Defensiveness, in turn, negatively predicted men's intentions to engage in action to support women's rights, and positively predicted their intentions to promote the rights of their own (advantaged) group.

Group interaction as a means of mitigating defensiveness and engaging men in solidarity-based action via the formation of new social identities

How can we reduce men's defensiveness and increase their engagement in efforts to address violence against women? Research examining the *intragroup* processes through which individuals come to participate in coordinated collective behaviour may provide some answers. This work draws on the framework of the social identity approach (Tajfel & Turner, 1979) to propose that *new* social norms (that is, the appropriate and desired attitudes and behaviours in a given context) can emerge interactively through group members' discussion and debate (Bongiorno et al., 2016; Postmes et al., 2005, Thomas & McGarty, 2009, Smith & Postmes, 2009; 2011; Smith et al., 2015). Group interaction can facilitate the development of

new *opinion-based* social identities grounded in common beliefs about how group members should feel and behave (McGarty et al., 2009). That is, while people can derive norms and attitudes from pre-existing social identities, there is also another possibility: That people can inductively generate norms and attitudes through interaction with like-minded others and, in doing so, develop *new* identities (Postmes et al., 2005). These new (or *emergent*) identities can provide the foundation for coordinated group action (Haslam et al., 1998; Smith et al., 2009; 2011; Thomas & McGarty, 2009; Thomas et al., 2016).

A considerable body of work has provided evidence for the role of group interaction in fostering people's engagement in different forms of social and political protest. Importantly, this work has demonstrated that group interaction can motivate both prosocial actions (e.g., solidarity-based action to support disadvantaged groups) and antisocial behaviour (e.g., outgroup hostility) through the development of opinion-based identities – depending on the particular norms discussed by the interacting group (Smith et al., 2009; 2011; Thomas & McGarty, 2009). For example, Thomas et al. (2016; 2019) demonstrated that group-based interaction regarding strategies to achieve clean water in developing nations boosted people's commitment to combat poverty via participants' identification with other supporters of the anti-poverty cause. Conversely, Smith and Postmes (2011; following Myers & Bishop, 1970) found that group interaction had mobilising effects on *hostile* forms of intergroup behaviour. When participants' discussions validated prejudicial stereotypes about the outgroup (immigrants in Britain), people reported greater intentions to engage in discriminatory behaviour by casting a vote for a fictitious anti-immigration political party (Smith & Postmes, 2011).

Taken together, this research highlights the power of group interaction in shaping people's attitudes, identity formation, and action intentions. Indeed, participation in coordinated, collective behaviour has always relied on people's interactions with one another.

Through airing grievances and discussing and validating opinions, people can come to a shared agreement regarding a desired course of action (Smith et al., 2015; Thomas & McGarty, 2009; Thomas et al., 2022). In everyday social interactions (both online and offline), women's disclosures of sexual harassment and assault elicit men's discussions and debate. The content of these discussions is likely to be important for their sense of who they are and how they intend to act. Yet, it remains to be tested whether targeted group interaction could provide an effective means of reducing defensive reactions and increasing their engagement in action to address the issue of violence against women.

The Current Research

The current chapter draws on the social identity approach (Tajfel & Turner, 1979; Turner et al., 1987) to examine the effects of group interaction on men's commitment to address gender inequality. In four pre-registered studies, I test whether group interaction centred on promoting justice for women provides the basis for the development of an opinion-based identity rooted in addressing the issue of sexual violence against women. Figure 8 anticipates my approach and hypotheses. I expect that when men are defining themselves in terms of an identity that values taking action to address gender inequality, defensiveness should be less likely to arise, which should, in turn, have positive flow on effects for men's intentions to engage in collective action for women's rights (Figure 8).

As in Chapter 3, in the present studies defensiveness was modelled as a latent variable indicated by three common ways that men can protect against threats to their group's morality: (a) by competing with women for victim status (*competitive victimhood*; Noor et al., 2012; Sullivan et al., 2012), (b) by perceiving men as experiencing *more* harm due to the issue of violence against women compared to women themselves (Bilali et al., 2012), and (c) by engaging in *pseudo self-forgiveness* (denying any ingroup wrongdoing and blaming women for the issue of gendered violence; Fisher & Exline, 2006; Hall & Fincham,

2005; Woodyatt & Wenzel, 2013). This latent approach allows for defensiveness to be captured by a range of indicators, and statistically it allows for testing complex hypotheses while accounting for measurement error.

In the studies presented within this chapter, men's collective action intentions were measured using items that included their engagement in both offline and online actions to promote gender equality (e.g., "I intend to attend a rally aimed at raising awareness of sexual harassment and assault against women", "I intend to raise awareness of women's experiences of sexual harassment and assault by posting on social media"). This measure of intentions was supplemented with an observed behavioural measure. Participants were presented with a link redirecting them to sign a petition designed to reauthorize legislation protecting victims of domestic and/or sexual violence. Whether or not participants clicked on this link was recorded using Qualtrics software.

Unlike previous research using face to face group interaction in the lab (e.g., see Thomas et al., 2009; 2016; 2019), the current studies use ChatPlat – an online chat tool compatible with Qualtrics software – to facilitate discussion between participants (Brooks, 2015). ChatPlat allows participants to interact anonymously via instant messages sent to one another in real time.
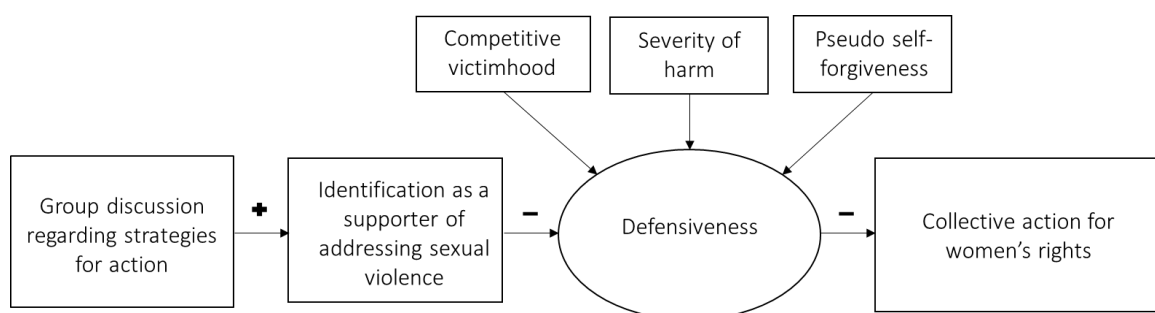

Figure 8. Conceptual model. Group discussion promotes increased collective action (intentions and behaviour) for gender equality via increased identification as a supporter of addressing sexual violence and reduced defensiveness.

Overview of Studies

In Study 5, participants engaged in a short online group interaction task centred upon discussing strategies to address violence against women and promote gender equality (relative to a comparative individual reflection task or control condition), before completing a questionnaire that included self-reported measures of the key dependent variables. The primary analyses use these self-reported quantitative responses of identification, defensiveness, and collective action (as well as the behavioural measure of action), to test the hypothesized model (Figure 8). The control condition was included in this first initial study to provide a baseline measure of the key outcomes.

In Study 6, I extend the online group interaction paradigm used in Study 5 to test whether group discussion has any lasting effects on men's social identification, defensiveness, and commitment to engage in collective action. I adopt a repeated measures design that surveys participants directly post-intervention and at two-week follow-up. In Study 7, I examine the content of participants' group discussions obtained from Studies 5 and 6 using natural language processing to make sense of the mixed findings. In this study I test the proposition that not all discussions will be mobilising (see Smith et al., 2018), and that group interaction will be more strongly associated with action outcomes when the discussions use language specifically related to collective action strategies.

In Study 8, in addition to the mode of reflection (group discussion vs. individual reflection), I manipulate the topic of reflection to experimentally test the influence of collective-action relevant language on supporter identification, defensiveness, and commitment to action. I again analyse the content of participants' discussions to examine differences in language use across experimental conditions. The methods, approach, and hypotheses for all three experimental studies were pre-registered. Links to the pre-registration

documents can be found in the reporting below; I note herein any deviations from the pre-registered approach.

Study 5

Study 5 adopts a between-subjects design with three conditions: group discussion, a comparative individual reflection task without group discussion, and a control group whereby individuals solely completed the dependent variables of interest[1]. All participants read information about the rates of sexual harassment and assault against women and girls. Participants in the group discussion and individual reflection tasks were then asked to brainstorm strategies to reduce the issue of gender-based violence. However, those in the group discussion condition completed the task in small online discussion groups, while those in the individual reflection completed it alone. The control group did not complete any task. I test my hypothesized conceptual model (Figure 8) by exploring whether (a) online group discussion regarding strategies to combat sexual violence against women increases men's support for action to address sexual harassment and assault against women; and (b) whether opinion-based identification (as a supporter of addressing sexual violence against women) and defensiveness mediate this effect.

Method

Participants. A pre-registration document for this study is available at: https://osf.io/bs4yd/?view_only=994b89c0e72744749be616ba366fa15c. A priori power analyses were conducted using Monte Carlo simulations for a two-level mediation model in Mplus (Muthén & Muthén, 2002). My main hypothesis requires testing an indirect effect of group discussion on collective action via opinion-based identification and defensiveness. I assumed a small effect size based on small – medium effects commonly reported in social

---

[1] Although I use the term "group interaction" elsewhere, in the Method, Results, and Discussion sections, I refer to this as "group discussion" to avoid confusion between the manipulations and reporting of statistical interaction.

psychology (β = .20 – β = .50; Cohen, 1988; Richard et al., 2003). Results revealed that 300 observations (group discussion, N = 25 groups of 4; individual reflection, N = 100; control, N = 100) would be required to detect an indirect effect with 80% probability at alpha = .05. Given the possibility that some participants may drop out of live chat during discussion, I anticipated that the average group size would be between 3-5 participants. To ensure that there were enough participants left to complete the task I initially set the maximum group size to 5, before reducing this to 4 due to delays in recruitment.

I recruited 314 participants from the United States who identified as men via Cloud Research (TurkPrime); average age = 40.33 (SD = 12.29). To maximize data quality, participants with an approval rating of at least 90% were invited to participate. Participants were excluded from analysis if technical difficulties prohibited them from participating in group discussion, leaving a total of 299 participants for analysis. This left 99 participants in the group discussion condition for analyses. The final configuration of chat groups (N = 28) included N = 17 groups of four, N = 9 groups of three, and N = 2 groups of two. Participants were predominantly White (83.3%), heterosexual (91.3%), and 65.9% held a bachelor's degree or higher. Political orientation was slightly left leaning (M = 3.46, SD = 1.79) on a 7-point scale ranging from 1 = *Strongly Liberal* to 7 = *Strongly Conservative*.

Procedure. The task requirements were similar across all three conditions, however participants in the group discussion condition completed the task interactively (in small online chat groups) whilst participants in the individual reflection condition completed the task alone. All participants accessed the study via a link on Amazon's CloudResearch and were told that the survey was interested in investigating people's attitudes towards a range of social and political issues. To provide context for the task, participants first read information regarding rates of sexual harassment and assault against women in America (Employee Equal Opportunity Commission, 2021; Rape, Abuse, & Incest National Network; 2021):

> Women and young girls are significantly more likely to be the victims of sexual harassment and assault. For instance:
>
> • In the United States, 1 in 4 women have experienced rape or attempted rape during their lifetime. 82% of all juvenile victims of sexual assault are female, while 90% of adult rape victims are female.
>
> • The United States Equal Employment Opportunity Commission has reported that women disproportionately report sexual harassment in the workplace, across all industries.

Participants in the group discussion and individual reflection conditions (but not the control condition) were told that the researchers were interested in ways that men can help to address sexual harassment and assault against women, and that their task was to come up with 5 practical strategies that men can implement to (a) promote gender equality, (b) raise awareness of women's experiences of sexual harassment and assault, and (c) reduce sexual harassment and assault against women. All participants were told that their strategies would be written up by the researchers and sent to organisations advocating for gender equality. Participants in the group interaction condition were then placed into a live online chat embedded within the Qualtrics survey with 3-4 other participants (ChatPlat, 2015). Participants – all identified by anonymous usernames and therefore unknown to one another – were able to engage in group discussion by sending instant messages for up to 15 minutes. The instructions regarding their task remained above the chat interface. They were notified two minutes before the chat ended and after 15 minutes the chat window closed automatically. Following this, participants completed the rest of the questionnaire.

Participants in the individual reflection condition did not participate in group discussion and were instead provided with a text box to record their strategies before proceeding to the remainder of the survey. Participants in the control condition did not participate in any task but instead proceeded to the questionnaire immediately after reading the information about rates of sexual violence and harassment against women in America.

After completing the key dependent measures and demographics, participants were debriefed and reimbursed USD$1.50 – 4.50 for their time (participants who engaged in group discussion were paid more due to the higher level of time and investment required).

Measures.

All items were measured on a 7-point Likert-type scales anchored 1 ('Strongly disagree') and 7 ('Strongly agree') unless otherwise specified.

Identification as a supporter of efforts to end violence against women. Social identification with a group defined by support for addressing sexual harassment and assault against women was assessed using 5 items adapted from Leach et al. (2008)'s measure of ingroup identification ($\alpha = .91$), e.g., "I am glad to be a supporter of efforts to end sexual harassment and assault against women", "The fact I am a supporter of efforts to end sexual harassment and assault against women is an important part of my identity".

Defensiveness. As in Chapter 3, defensiveness was operationalised as a latent variable with three reflective indicators: competitive victimhood (Noor et al., 2008; 2012), minimization of the severity of harm (Bilali et al., 2012), and pseudo self-forgiveness (Hall & Fincham, 2005; Woodyatt & Wenzel, 2013). Specifically, four items adapted from Kahalon et al. (2019)'s measure of competitive victimhood measured the extent to which men competed with women for victim status, e.g., "Economically, men in America are discriminated against more than women", "Men in America are now suffering more emotional pain than women", $\alpha = .94$.

Participant's perceptions of the *severity of harm* inflicted on their own group (men) compared to women due to sexual harassment in the workplace were captured using one item adapted from Bilali et al. 2012, "Which group experiences *more* harm as a result of sexual harassment against women in the workplace?". Participants indicated their response on a

bipolar scale from 1 = *women* and 7 = *men,* such that higher scores reflected the perception that men suffer *more* harm due to sexual harassment compared to women.

Finally, three items measured the extent to which men engaged in the process of *pseudo self-forgiveness* (Hall & Fincham, 2005; Woodyatt & Wenzel, 2013) by denying their group's involvement and deflecting blame, e.g., "I'm not really sure whether what men did was wrong", "Men aren't the only ones to blame for what happened", α = .74.

Collective action intentions. The extent to which participants were committed to addressing sexual harassment and assault against women was measured using five items that sought to capture men's intentions to engage in a range of online and offline actions, e.g., "I intend to…raise awareness of women's experiences of sexual harassment and assault by posting on social media", "…sign a petition advocating for the end of sexual harassment and assault against women", "…attend a rally aimed at raising awareness of sexual harassment and assault against women", "…encourage others to support ending sexual harassment and assault against women". These items were aggregated together to form an internally consistent scale (α = .90).

Behavioural measure of prosocial action. A single measure of observed behaviour was taken by capturing whether or not participants clicked on a link relevant to expressing support for legislative change designed to protect survivors of domestic and sexual violence and their families (clicked on the link: coded '1', did not click on the link: coded '0').

<div align="center">

Results and Discussion

</div>

Analytic Strategy

Given that participants engaged in group discussion in one of the conditions in all three experimental studies (Study 5, 6 and 8), the data involve a combination of independent and statistically non-independent observations (that is, responses from individuals in the group discussion condition are nested within interacting groups). Table 13 displays the

intraclass correlations for the key variables for participants in the group discussion condition. This table shows that between 5-31% of the between groups variance in those variables is due to the interactive group that participants were assigned to. Accordingly, my analysis adopts a multi-level approach by clustering the data at the group level to control for the fact that the full dataset involves responses from individuals (the questionnaire responses; a within persons factor) that are nested within discussion groups. In Study 5, 6, and 8, these self-reported questionnaire responses are analysed using multi-level SEM in Mplus Version 8 to control for nonindependence (Muthén & Muthén, 2017). The path model (Figure 8) is specified at the within-person level of analysis – that is, I only consider the relationships at the individual level (the within-persons, self-reported questionnaire responses). I include the standardized root mean squared residual (SRMR) as an indicator of model fit as it is considered to be the only reliable indicator of fit for multilevel models (Muthén & Muthén, 2007). A value of .08 or less is generally considered good fit for SRMR (Hu & Bentler, 1999).

Preliminary analysis: The effects of group discussion

The primary focus in Study 5 was to provide an initial test of my hypothesized model (Figure 8) to establish whether group interaction increased collective action intentions and behaviour (relative to individual reflection) via identification as a supporter of efforts to end sexual harassment and assault and defensiveness. I first used the full dataset to map out the direct effects of completing the reflection task individually or in interactive groups (as well as making comparisons with the control group). I created effect codes to compare the influence of the group discussion, individual reflection, or control conditions in Mplus using maximum likelihood parameter estimates with a Huber-White (Sandwich) estimator to correct for non-independence of observations (MLR). Table 13 displays the intraclass correlation coefficients

and the intercorrelations for the key outcome variables, as well as the means, standard deviations, and confidence intervals for all three conditions.

I first sought to establish whether individually reflecting on strategies to address violence against women influenced participants' social identification as a supporter, defensiveness, and collective action intentions and behaviour – relative to the passive control condition. That is, does individual reflection on practical strategies stimulate engagement relative to a passive control? Table 13 shows that there were no significant differences in the key dependent variables between participants in the individual reflection condition and control condition (all t's = -1.09-1.10 and all p's = .27-.80). These results suggest that individually reflecting on the issue of gender-based violence did not have any influence on the key outcomes over and above the passive control condition. Accordingly, in my subsequent analyses I used the individual reflection condition as the primary comparison group.

I next sought to establish the effects of engaging in group discussion on the key dependent variables. Table 13 shows that, relative to individual reflection, group discussion had direct effects on identification as a supporter (β = .16, p <.001), collective action intentions (β = .13, p = .01) , and pseudo self-forgiveness (β = -.13 and p = .01) but did not affect attributions of harm (β = -.03, p = .70), competitive victimhood (β = .00, p = .98), or the behavioural measure of collection action (β = -.01 , p = .37).

Table 13. Intra-class correlation coefficients, mean individual-level scores, standard deviations, 95% confidence intervals, and inter-scale correlations across the group discussion, individual reflection, and control conditions between the key variables of interest.

Variable	Intra-class correlation coefficient	Group discussion ($N = 99$)	Individual reflection ($N = 101$)	Control ($N = 99$)	1	2	3	4	5	6
1. Identification*	.31	5.49ª (1.00) [5.29, 5.70]	5.13ᵇ (1.28) [4.88, 5.38]	5.00ᵇ (1.40) [4.72, 5.28]	-					
2. Competitive victimhood	.27	2.63ª (1.66) [2.30, 2.96]	2.62ª (1.56) [2.31, 2.93]	2.80ª (1.61) [2.47, 3.11]	-.36**	-				
3. Pseudo self-forgiveness	0.14	2.47ª (1.25) [2.22, 2.72]	2.79ᵇ (1.14) [2.56, 3.01]	2.97ᵇ (1.19) [2.73, 3.21]	-.43**	.55**	-			
4. Severity of harm	0.23	1.70ª (1.41) [1.42, 1.98]	1.77ª (1.30) [1.52, 2.03]	1.93ª (1.60) [1.61, 2.25]	-.24**	.36**	.36**	-		
5. Collective action intentions	0.24	4.69ª (1.48) [4.39, 4.98]	4.27ᵇ (1.67) [3.94, 4.60]	4.00ᵇ (1.87) [3.62, 4.37]	.67**	-.36**	-.41**	-.15*	-	
6. Clicked behavioural measure (%)	0.05	12.1%ª	14.9%ª	16.2%ª	.15**	-.13*	-.10	.00	.24**	-

Note. ᵃ, ᵇ Within a row, means without a common subscript differ ($p < .05$).
**$p < .001$. *$p < .05$.
*Identification as a supporter to address sexual harassment and assault against women

Testing the hypothesized model. To test my hypothesized mediation model, I conducted a multilevel structural equation model (MSEM; Bovaird, 2007) in Mplus (version 8.1; Muthén & Muthén, 2017). The Mplus command TYPE = COMPLEX was used to account for the nested nature of the data (Hox, 2002).

I first tested a mediation model whereby group discussion (dummy coded as '1') was a direct predictor of identification as a supporter of efforts to address sexual harassment and assault against women. The individual reflection condition was dummy coded as the reference group ('-1'). Identification as a supporter of addressing sexual violence predicted defensiveness (latent variable), which, in turn, predicted collective action intentions. This model fitted the data well, SRMR = .07. Figure 9 shows that all paths were significant (values to the left of the forward slash). In line with my predictions, the indirect effect of group interaction on action intentions via identification and defensiveness was significant, IE = .08, 95% CI [0.01, 0.14].

I next tested the same model using the dichotomous behavioural measure of collective action as the outcome variable. This model also had good fit with the data (SRMR = .06). Similar to the measure of collective action intentions, all paths were significant (Figure 9, values to the right of the forward slash). As predicted, there was a significant indirect effect of group discussion on behavioural action via identification as a supporter and defensiveness, IE = .01, 95% CI [0.00, 0.02].
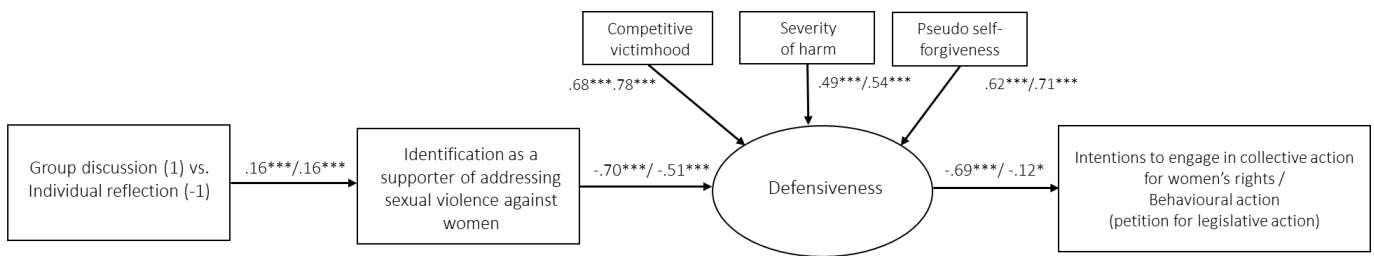
Figure 9. Indirect effect of group discussion on collective action intentions and behaviour via identification as a supporter of addressing sexual violence against women and defensiveness. Values obtained when collective action intentions was modelled as the outcome are to the left of the forward slash, values for the behavioural measure of collective action are to the right. Values are standardized coefficients (*p < .01; ***p < .001).

The results of this first study reveal that group discussion (relative to individual reflection) was positively associated with men's collective action intentions and behavioural action to address sexual harassment and violence against women. In line with my predictions, the effect of group discussion on collective action was mediated by men's increased identification as a supporter of efforts to address sexual harassment and assault against women, and reduced defensiveness regarding the issue of gender-based violence. Accordingly, the results suggest that men's participation in group discussion regarding strategies to address gender inequality uniquely contributed to the formation of a social identity centred upon addressing sexual violence against women. This identity, in turn, influenced men's collective action intentions and behaviour by reducing defensiveness (competitive victimhood, minimization of harm, and pseudo self-forgiveness) regarding the issue of sexual violence against women.

These findings provide strong initial evidence regarding the transformative influence of group discussion on men's social identification, defensiveness, and collective behaviour.

However, it is unclear whether group discussion has lasting effects on men's identity, defensiveness, and action over time. In Study 6 I therefore sought to test the robustness of the effects found in Study 5 and test whether group discussion has any sustained effects on men's collective action. In Study 6 I took measures of the main dependent variables post-intervention (group discussion vs. individual reflection) as well as at two-week follow-up. Given there were no differences in key outcomes between the individual reflection condition and the control condition in Study 5, Study 6 did not include a control condition and participants were allocated to either group discussion or individual reflection.

Study 6

Method

Participants. Study 6 used a 2 x 2 (intervention: group interaction, individual reflection; time: post-intervention, two-week follow-up) mixed-subjects design. See here for the pre-registration documentation: https://osf.io/dk8wc/?view_only=1e8f1bfd90154b7bba4c84c5361edc5b. Participants ($N = 248$) were again recruited via Amazon's CloudResearch ($M^{age} = 39.92$; $SD = 11.87$). Five participants were removed due to failing an initial bot screening test and three were removed due to technical issues precluding them from participating in the group interaction task. This left 240 participants at Time 1 ($N = 118$ in the group discussion condition; $N = 122$ in the individual reflection condition). At two-week follow-up, $N = 107$ participants from the group interaction condition and $N = 114$ from the individual reflection condition completed the post-intervention questionnaire. In Study 2, participants were predominantly White (76%), heterosexual (90.4%), and 58.7% held a bachelor's degree or higher. Political orientation of the sample was slightly left leaning where 1 = *strongly liberal* and 7 = *strongly conservative* ($M = 3.41$, $SD = 1.77$).

Procedure. The first part of this study used the same procedure as Study 5; however, participants were notified that the study involved participation at two timepoints before indicating their agreement to participate. As in Study 5, participants in the group discussion condition were allocated to groups of four using ChatPlat software embedded within the Qualtrics survey, while participants in the individual reflection condition completed the reflection task using the text box provided. Following this, participants completed the self-reported key dependent measures and demographic variables used in Study 5 (all measures formed reliable scales, α_{range} = .70–.94). After two weeks, participants were contacted using CloudResearch's MTurk Toolkit and provided with the link to a Qualtrics survey that contained the same dependent measures. Following completion of this survey, participants were thanked, debriefed, and reimbursed between USD 3.00-$4.00 for their participation.

In this study there was an administrative error during data collection such that the variable that linked participants in the group discussion condition to the particular group they were assigned to was not captured. However, in the analyses outlined below I used a proximal measure of the clustering variable by triangulating the time at which participants entered the online chat to cluster the data at the group level. I consider the implications of the error and limitations of this approach in the general discussion section.

Results and Discussion

Preliminary analysis: The effects of group discussion

The aim of Study 6 was to replicate the findings of Study 5 and test if there was any effect of group discussion on participants' collective action intentions and behaviour at two-week follow-up. Table 14 includes the intraclass correlation coefficients for the key outcomes across time points, which shows that between 1-5% of the variance in participants' responses was due to the interactive group to which they were assigned (1-3% at Time 2). I therefore created effect codes in Mplus (group discussion = '1', individual reflection = '-1') using

MLR with Huber-White (Sandwich) estimator to correct for the non-independent nature of the data (as in Study 5).

Table 14. Intra-class correlation coefficients at Time 1 and 2.

Variable	Time 1	Time 2
Identification as a supporter	0.01	0.03
Competitive victimhood	0.04	0.02
Pseudo self-forgiveness	0.01	0.02
Severity of harm	0.02	0.03
Collective action intentions	0.05	0.03
Clicked behavioural measure (%)	0.03	0.01

Table 15 displays the means, standard deviations, and 95% confidence intervals for the group discussion and individual reflection condition at each time point. Intercorrelations between variables across time points can be found in Table 16. Contrary to expectations, Table 15 shows that engaging in group discussion (relative to individual reflection) had no direct effects on the key outcomes at both Time 1 (all β's = -.01-.09 and all p's = .14 -.63) and Time 2 (all β's = -.03-.09 and all p's = .19-.75).

Table 15. Frequency of petition clicks, mean individual-level scores (and standard deviations) between conditions for Time 1 and 2.

Note. *Identification as a supporter of efforts to address violence against women.

	Time 1		Time 2	
	Group interaction (N = 118)	Individual reflection (N = 122)	Group interaction (N = 107)	Individual reflection (N = 114)
Identification*	5.36 (1.23)[a] [5.12, 5.55]	5.42 (1.03)[a] [5.23, 5.59]	5.12 (1.19)[a] [4.90, 5.35]	5.27 (1.07)[a] [5.08, 5.47]
Competitive victimhood	2.81 (1.62)[a] [2.52, 3.11]	2.62 (1.53)[a] [2.36, 2.89]	2.84 (1.71)[a] [2.52, 3.17]	2.67 (1.47)[a] [2.42, 2.93]
Pseudo self-forgiveness	3.84 (0.76)[a] [3.70, 3.98]	3.86 (0.71)[a] [3.74, 3.99]	3.16 (1.27)[a] [2.93, 3.39]	3.03 (1.87)[a] [2.81, 3.25]
Severity of harm	1.73 (1.29)[a] [1.52, 1.97]	1.62 (1.20)[a] [1.43, 1.84]	2.04 (1.52)[a] [1.24, 1.77]	2.11 (1.68)[a] [1.81, 2.43]
Collective action intentions	4.89 (1.66)[a] [4.60, 5.16]	4.73 (1.43)[a] [4.47, 4.97]	4.34 (1.67)[a] [4.02, 4.65]	4.55 (1.43)[a] [4.29, 4.82]
Behavioural measure (legislative action) (%)	11.1 %	17.5 %	8.4%	14.0%

Table 16. Intercorrelations between key variables across Time 1 and 2.

	1	2	3	4	5	6
1. Identification*	-					
Time 1	-					
Time 2	-					
2. Competitive victimhood		-				
Time 1	-.40**	-				
Time 2	-.36**	-				
3. Pseudo self-forgiveness			-			
Time 1	-.41**	.68**	-			
Time 2	-.25**	.46**	-			
4. Severity of harm				-		
Time 1	-.30**	.50**	.47**	-		
Time 2	-.33**	.36**	.57**	-		
5. Collective action (CA) intentions					-	
Time 1	.69**	-.25**	-.34**	-.22**	-	
Time 2	.70**	-.27**	-.19**	-.26*	-	
6. Behavioural measure						-
Time 1	.15*	-.19**	-.08	.17*	.17**	-
Time 2	.12	-.10	-.07	-.09	.19*	-

Note. $*p < .05.$ $**p < .001.$
*Identification as a supporter of efforts to address violence against women.

Testing the hypothesized model.

As in Study 5, path analysis was conducted in Mplus Version 8 using TYPE = COMPLEX to control for the non-independence of data (i.e., individuals nested within interacting groups; Muthén & Muthén, 2017). Group discussion (dummy coded '1', individual reflection coded '-1') predicted identification as a supporter of efforts to address gendered violence. Identification predicted latent defensiveness, which, in turn, predicted collective action. I ran two mediation models at Time 1 using both collective action outcomes (intentions and behaviour), before running the models at Time 2 to examine if the effects held across time.

Contrary to expectations, at Time 1 (directly post-intervention) there was no significant direct effect of engaging in group interaction on identification as a supporter to address sexual harassment and assault against women (see Figure 10). This model did not fit the data well (SRMR = .11). The modification indices suggested that additional variance between opinion-based identification and collective action intentions needed to be accounted for. Adding this path significantly improved model fit (SRMR = .03). As expected, identification negatively predicted defensiveness, and positively predicted collective action intentions. However, there was no association between defensiveness and collective action intentions, suggesting that the variance was being usurped by the strength of the relationship between identification and collective action (Figure 10, values to the left of the slash). There was no significant indirect effect of group discussion on collective action through identification and defensiveness, IE = -0.00, 95% CI [-0.01, 0.00].

When the behavioural measure of action was modelled as the outcome, a similar pattern of results were obtained – however, defensiveness negatively predicted behavioural action (see Figure 10, values to the right of the slash; SRMR = .03). The indirect effect of group discussion on behavioural action via identification and defensiveness was also not significant, IE = -0.01, 95% CI [-.01, 0.01].

Figure 10. Results of path model at Time 1, Study 6. Values obtained when collective action intentions was modelled as the outcome are to the left of the oblique, values for the behavioural measure of collective action are on the right. Values are standardized coefficients

(***p < .001). Dotted lines represent non-significant paths. N.B. This additional pathway

from was only modelled from identification to collective action intentions (due to initial poor

model fit).

Similarly, at Time 2, group discussion had no significant effect on opinion-based

identification (see Figure 11; SRMR = .08). Identification negatively predicted defensiveness,

which was negatively associated with collective action intentions (Figure 11, values to the

left of the slash). There was no significant indirect effect of group discussion on collective

action through identification and defensiveness, IE = -0.03, 95% CI [-0.03, 0.01]. A similar

pattern of results was obtained when the behavioural measure of action was modelled as the

outcome (see Figure 11, values to the right of the slash; SRMR = .08). The indirect effect of

group discussion on behavioural action via identification and defensiveness also contained

zero and was not significant, IE = -0.01, 95% CI [-.01, 0.01].
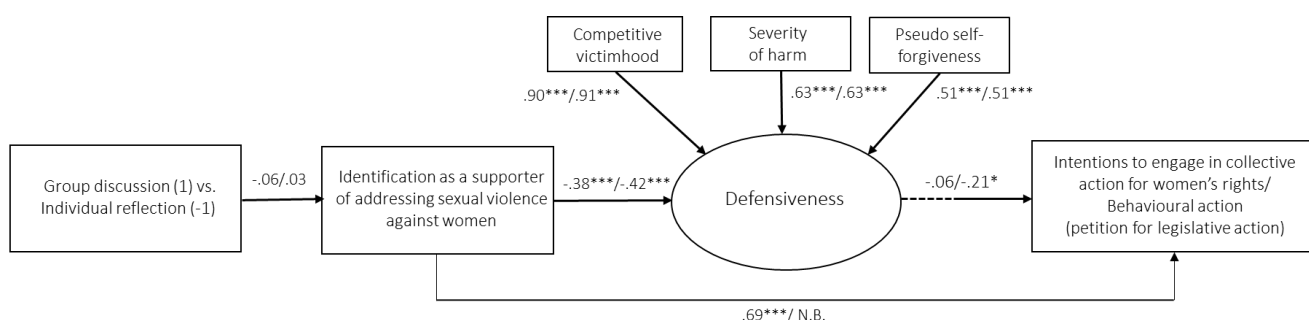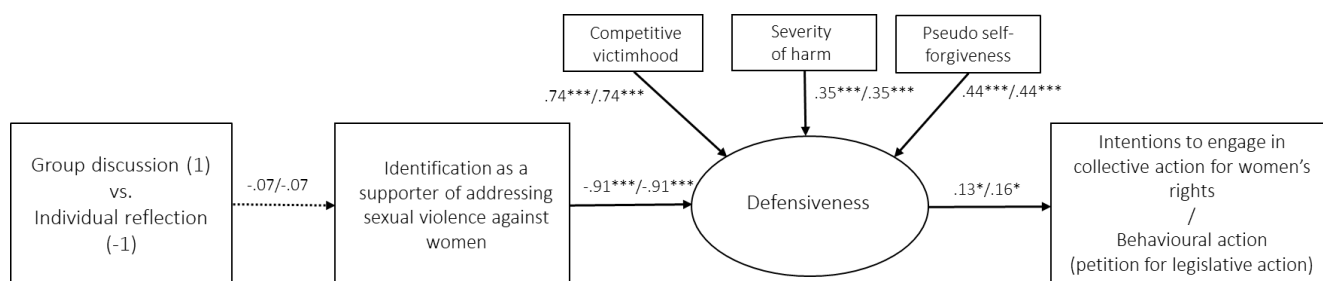

Figure 11. Results of path model at Time 2, Study 6. Values obtained when collective action

intentions was modelled as the outcome are to the left of the oblique, values for the

behavioural measure of collective action are on the right. Values are standardized coefficients

(***p < .001). Dotted lines represent non-significant paths.

In contrast to Study 5, the results of this study indicate that group discussion had little

direct or indirect influence on men's identification as a supporter of addressing sexual

violence, defensiveness, or collective action intentions and behaviour directly after the intervention and at two-week follow-up.

It is possible that these contradictory findings are a result of key features of the interacting groups that differed between Studies 5 and 6. Indeed, previous research has emphasized that the *content* of people's interactions is critical in facilitating the development of identities that are either prosocial or antisocial in nature (Thomas & McGarty, 2009; Thomas et al., 2016; Smith & Postmes, 2011). When interactions with others centre actions designed to address injustice and inequality, people can come to see themselves in terms of identities grounded in support for the fight to address inequality (Thomas & McGarty, 2009). In contrast, when interactions suggest that outgroup hostility is the desired course of action for the group, people are more likely to see themselves and others in terms of an identity based on negative attitudes towards the outgroup (Smith & Postmes, 2011).

Taken together, these findings suggest that the effectiveness of group discussion in facilitating the development of a *pro* gender equality identity is likely to hinge on the extent to which participants' discussions are focused on strategies designed to promote gender equality. To explore this possibility, in Study 7 I examined the content of the interactions (i.e., the chat transcripts) between Studies 5 and 6. I used natural language processing software to identify whether disparities in language use in the interacting groups could explain differences in the effects of group discussion across studies.

Study 7

To better understand the role of the content of group discussions in Studies 5 and 6, I conducted computerized text analysis on the pooled transcripts ($N = 59$) of the group discussions from Study 5 ($N = 28$) and Study 6 ($N = 31$) using Linguistic Inquiry Word Count analysis software (LIWC2015; Pennebaker et al., 2015). LIWC analyses written text using dictionaries that represent both linguistic categories (e.g., pronouns) and emotional and

cognitive components (e.g., negative affect), and calculates the proportion of words in each passage of text that match these predefined dictionaries (Pennebaker et al., 2015). Running the chat transcripts through LIWC allowed me to create quantitative variables to capture differences in language expression in the group discussions between Study 5 and 6. I expected that the effectiveness of group discussion in Study 5 (compared to Study 6) may have been due to the fact that, on average, participants' discussions in Study 5 used more collective action relevant language than the interactions between participants in Study 6.

Method

    Measures. I drew on LIWC's internal and custom dictionaries to identify (a) expressions indicative of an emergent social identity, (b) communication regarding illegitimate harmdoing (the acknowledgement of sexual violence as unjust) as a proxy for defensiveness, and (c) the expression of language related to engagement in collective action.

    Social identification. Based on previous studies using LIWC to identify social psychological variables (Vergani & Bliuc, 2020; Pennebaker et al., 2015; Smith et al., 2015; 2018), I used LIWC's internal dictionary subcategories of first- and third-person plural pronouns and words (e.g., we, us, they) and affiliation category (e.g., ally, friend, social) to capture social identification with others who value acting to address sexual violence against women. The mean proportions from these two categories were combined to create a single measure of social identification.

    Moral engagement. Graham et al. (2009)'s moral foundations dictionary (MFD 2.0) was used to index the proportion of words from each group relating to participants' acknowledgement of ingroup harm. I used this variable to map onto the latent defensiveness variable by coding for an *absence* of ingroup defensiveness. In the social psychological literature, the term *defensiveness* has been used to encompass a range of strategies individuals can take to defend against threats to their personal or group identity. One common defensive

strategy has been to *morally disengage* from the wrongdoing (Bandura 1990; 1996; 1999). *Moral disengagement* refers to the mechanisms people employ to avoid the negative consequences associated with their engagement in immoral behaviour (e.g., the experience of negative emotions such as shame and guilt) and is typically reflected by strategies that aim to deny or minimize the perpetrator group's role in causing harm (Bandura, 2011). Given this, I use the term *moral engagement* here to refer to the acknowledgement of ingroup harm (that is, an absence of moral disengagement strategies reflected in participants' acknowledgement of atrocities committed by their own group; Vollhardt et al., 2014). The MFD captures moral language using five different dimensions: care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and sanctity/degradation (Haidt & Graham, 2007; Graham et al., 2011; 2013). For each dimension, two words exist (a positive "virtue" and negative "vice"). I included the dimensions relating to care/harm (compassion towards victims and anger at injustice, harm, and individual suffering, e.g., "compassion, care, protect") and fairness/cheating (the desire for trustworthy systems and people, e.g., "rights", "equal", "justice", "unfair") as these categories contain words most related to the acknowledgement of harm and injustice.

Collective action. I used Smith et al.'s (2018) custom collective action dictionary to capture the proportion of words per group relating to intended engagement in collective action (e.g., "protest", "campaign", "volunteer").

Results and Discussion

Table 17 displays the mean proportion of words (and standard deviations) used in the interacting groups from the relevant LIWC categories across the two study samples. I conducted regression analyses while controlling for word count to assess whether the prevalence of the key variables differed across the two studies. Table 17 shows that there

were no differences in language related to social identification or moral engagement (discussions regarding harm and a desire for justice). However, there was greater use of collective action relevant language in the interacting groups in Study 5 compared to Study 6 (Table 17). That is, at the study level, the online discussions between participants in Study 5 compared to Study 6 differed in the amount that participants used language related to engagement in collective action. This finding may help to explain why the effects observed in Study 5 did not replicate in Study 6.

Table 17. Means and Standard Deviations for LIWC2015 Dimensions across Studies 5 and 6 samples.

LIWC* dimension	Study 5 ($n = 28$)		Study 6 ($n = 31$)	
	M	SD	M	SD
Word count	510.85[a]	169.55	536.90[a]	207.28
Identification	13.80[a]	4.78	12.50[a]	3.65
Moral engagement	1.30[a]	0.55	1.47[a]	0.86
Collective action	1.47[a]	0.91	1.10[b]	0.62

Note. [a] Within a row, means without a common subscript differ at $p \leq .05$.
*LIWC2015 = Linguistic Inquiry and Word Count (Pennebaker et al., 2015).

Given this, in Study 8 I directly manipulated the use of collective action language in order to experimentally test the hypothesis that it is not just any discussion that will be associated with identification as a supporter of efforts to address violence against women, reduced defensiveness, and commitment to action. Rather, discussions must centre on practical collective action strategies to address sexual harassment and assault, and not just attitudinal support per se, in order to reliably influence action intentions and behaviour.

Study 8

The results from Study 5, 6, and 7 suggest that the effectiveness of group discussion on men's commitment to act for gender equality is contingent on the extent to which

participants' discussions are focused on practical collective action strategies designed to address gender inequality. Thus, in Study 8 I sought to experimentally test the importance of collective-action relevant language by manipulating the topic of participants' reflection task – as well as whether or not they completed this task individually or in groups. The topic of reflection was manipulated via instructions that asked participants to come up with actions that men can take to address sexual harassment and assault against women (as in Studies 5 and 6), compared to a control condition whereby participants were instructed to brainstorm about why the *issue* of sexual harassment was important (but not any strategies to address it). I reasoned that for group discussion to have any mobilising effects on action, the content of participants' interactions must be concentrated on sharing concrete strategies for action – rather than simply talking about supporting the issue.

I hypothesized that (a) group discussion would lead to increased collective action intentions (and behavioural action, captured by whether or not participants clicked on a petition link relevant to expressing support for the Violence Against Women Act) relative to individual reflection. I expected that this would occur when participants discussed action-related strategies compared to when they solely discussed the issue of sexual harassment and assault (that is, I expected an interaction between mode of reflection and topic of reflection); and (b) that this effect would be mediated by increased identification as a supporter of addressing sexual violence. Identification should, in turn, be negatively associated with engagement in defensive strategies (competitive victimhood, minimization of harm, and pseudo self-forgiveness). Thus, my primary hypothesis was one of moderated mediation as I expected that the effect of group interaction on collective action via identity and defensiveness would occur for participants who discussed strategies to address violence against women, but not for those who solely discussed their opinions on the issue.

Method

  Participants and design. Study 8 used a 2 (mode of reflection: group discussion vs. individual reflection) x 2 (topic of reflection: action strategies vs. control) between-subjects design. As in the previous two studies, respondents ($N = 244$) were recruited through Amazon's Mechanical Turk (mean age = 38.82), sampling for participants who identified as men. 88% of participants identified as heterosexual. 73% of participants identified as White, 9.4% as Black, 9% as Hispanic, 6.1% as Asian and 1.2% as multi-racial. 66% of participants held a bachelor's degree or higher. Political orientation in this study was again slightly left leaning on a 7-point scale ($M = 3.47$, $SD = 1.77$).

  Within the group discussion condition, participants were allocated to groups of three or four on ChatPlat. Those who were unable to participate in the group chat portion of the study due to technical difficulties were removed, leaving 25 discussion groups ($N = 73$ participants) in the action strategy condition and 24 ($N = 73$) in the control condition. 96 participants completed the reflection task individually (action strategies, $N = 42$; control, $N = 54$).

  Procedure. In the action strategies group discussion condition, the procedure was identical to the group discussion condition in the previous two studies (i.e., participants were asked to undertake a 15-minute task using an online chat tool to discuss strategies designed to combat sexual harassment and assault). Specifically, participants were instructed to come up with five practical action strategies (i.e., things that men can go out and do) to 1) raise awareness of women's experiences of sexual harassment and assault, 2) reduce sexual harassment and assault against women, and 3) promote gender equality. In the control instructions group discussion condition, participants were asked to undertake group discussion regarding why the *issue* of sexual violence against women was important: "In this study we are interested in investigating men's opinions about sexual harassment and assault

against women. As a group, we ask that you take 15 minutes to discuss and come up with 5 reasons why sexual assault and harassment against women is an important issue." Participants allocated to complete the task individually were allocated to receive either action strategies instructions or control instructions (i.e., "On your own, we ask that you…"). These participants recorded their responses using the text box provided. Upon completion of the questionnaire, all participants were debriefed and compensated with USD [1.00-3.00] for their time.

The key (self-reported) measures included in the Qualtrics survey remained the same as Study 5 and 6 (α_{range} = .65 – .94). To check whether the instructions successfully manipulated the content of participants' reflections, responses were coded using Smith et al.'s (2018) custom LIWC collective action dictionary to create an observed variable of collective action relevant language (as in Study 7).

Results and Discussion

Preliminary analysis: The direct and interactive effects of topic of reflection and mode of reflection

Given the data obtained from the LIWC content analysis were independent (each interacting group had just one set of scores), I first conducted an independent samples t-test in SPSS using the observed measure of collective action language to assess whether the topic of reflection instructions successfully manipulated the content of participants' responses. The manipulation was successful: there was significantly greater collective action language used by participants in the action strategies instructions condition (hereafter referred to as the 'action condition', M = 1.42, SD = 1.12), compared to participants in the control instructions condition (M = 0.82, SD = 0.96), $t(229.03)$ = -4.57, p <.001, d = 0.60.

Table 18. Intra-class correlation coefficients (Study 8).

Variable	Intra-class correlation coefficient
Identification as a supporter	0.18
Competitive victimhood	0.10
Pseudo self-forgiveness	0.17
Severity of harm	0.07
Collective action intentions	0.17
Clicked behavioural measure (%)	0.05

As in Studies 5 and 6, I next used the full dataset to map out the direct and interactive effects of completing the task individually or in interactive groups and the topic of reflection. Table 18 shows that between 5-18% of the between groups variance in the dependent variables was due to the interactive group that participants were assigned to. Thus, I created effect codes to compare the influence of participating in group discussion (1) vs. individual reflection (-1) and action instructions (1) vs control instructions (-1) on the key outcomes in Mplus using MLR with Huber-White (Sandwich) estimator to correct for non-independence of observations. Descriptive statistics for the individual- (the self-reported questionnaire responses; a within persons factor) and group-level (observed content of reflection; a between-persons factor) variables are reported in Table 19, as well as the intercorrelations between key variables.

Table 19 suggests that the manipulations had little influence on the key dependent variables. Initial tests revealed that the main effect of topic of reflection on identification as a supporter to end sexual harassment and violence against women was not significant, $\beta = 0.00$, $p = .10$. Contrary to predictions, the two-way interaction between topic of reflection and mode of reflection on opinion-based identification was not significant, $\beta = 0.07$, $p = .33$. However, there was a marginal effect of mode of reflection (group discussion vs. individual reflection) on identification, $\beta = 0.11$, $p = .09$. Further, Table 19 includes 95% confidence intervals around the mean for participants in the four experimental conditions. Given that

responses in the individual reflection condition are statistically independent (participants

responses are not nested within interacting groups) these confidence intervals can be used to

infer differences between means in other cells. Table 19 shows that the mean for

identification in the group discussion x action condition lies outside the confidence intervals

for the individual reflection conditions. Notwithstanding the test of the overall interaction

term, this suggests that group interaction (for participants who discussed action strategies) did

in fact produce an increase in identification as a supporter on average when compared to

participants in the individual reflection conditions (action strategies and control instructions).

There was a significant main effect of mode of reflection (group discussion vs.

individual reflection) on the behavioural measure of collective action, $\beta = 0.13$, $p = .002$.

Neither main effects nor two-way interactions were significant for the three defensiveness

outcome measures or collective action intentions (all p's > .13). However, examining Table

19 shows that the collective action intentions mean in the group discussion x action condition

lies above the confidence intervals in the individual reflection conditions. Again, this result

suggests that despite the overall test of the interaction term, group interaction (for participants

who discussed action strategies) did result in greater collective action intentions on average,

compared to participants who engaged in individual reflection.

Table 19. Frequency of petition clicks, mean individual-level and group-level scores, and scale intercorrelations, in the action discussion (*N* = 74), control discussion (*N* = 73), action reflection (*N* = 42), and control reflection (*N* = 54) conditions, Study 8.

Individual level variables	Group discussion		Individual reflection		1	2	3	4	5	6	7
	Action instructions (*N* = 74)	Control Instructions (*N* = 73)	Action instructions (*N* = 42)	Control Instructions (*N* = 54)							
1. Identification as a supporter of addressing violence against women	5.47[a] (1.16) [5.19, 5.74]	5.34[a] (1.22) [5.06, 5.62]	4.94[a] (1.31) [4.57, 5.31]	5.16[a] (1.22) [4.83, 5.48]	-						
2. Competitive victimhood	2.84[a] (1.61) [2.46, 3.22]	2.59[a] (1.54) [2.21, 2.98]	2.98[a] (1.84) [2.48, 3.49]	3.02[a] (1.74) [2.58, 3.47]	-.42**	-					
3. Pseudo self-forgiveness	2.97[a] (1.20) [2.69, 3.24]	2.84[a] (1.18) [2.57, 3.12]	3.01[a] (1.20) [2.64, 3.37]	3.07[a] (1.20) [2.75, 3.40]	-.41**	.65**	-				
4. Severity of harm	1.89[a] (1.49) [1.56, 2.23]	1.88[a] (1.42) [1.54, 2.22]	2.12[a] (1.69) [1.67, 2.57]	1.81[a] (1.38) [1.42, 2.21]	-.35**	.50**	.51**	-			
5. Collective action intentions	4.93[a] (1.60) [4.56, 5.29]	4.37[a] (1.64) [4.01, 4.74]	4.19[a] (1.84) [3.71, 4.67]	4.36[a] (1.38) [3.91, 4.78]	.63**	-.41**	-.36**	-.27*	-		
6. Behavioural measure (petition for legislative change) (%)	20%[a] [12.2%, 27.8%]	13.7%[a] [5.8%, 21.6%]	11.9%[b] [1.5%, 22.3]	5.6%[b] [-3.6%, 14.7%]	.17**	-.21*	-.19**	-.06	.27**	-	
Group level variables											
7. Collective action language (LIWC)	1.39[a] (0.98) [1.15, 1.63]	0.76[b] (0.61) [0.52, 0.10]	1.50[a] (1.34) [1.19, 1.82]	0.90[b] (1.29) [0.62, 1.18]	.01	.01	.00	.05	.10	.08	-

Note. [a] Within a row, means without a common subscript differ at *p* <.05.

Testing the hypothesized model.

Testing the hypothesized model involved running two separate mediation models at different levels of the moderator (i.e., action instructions, 1 vs. control instructions, -1) for both collective action outcomes (intentions and behaviour) using Mplus v. 8 (Muthén & Muthén, 2017). This approach allowed me to test whether the indirect relationship between group interaction and collective action (intentions and behaviour) via identification and defensiveness differed depending on the topic of participants' reflection task.

Collective action intentions. I expected that group discussion would have an effect on collective action (intentions and behaviour) via identification and defensiveness when participants were instructed to brainstorm concrete collective action strategies to address gender-based violence (action instructions = 1). This model evidenced acceptable (although borderline) fit with the data, SRMR = .09. Consistent with predictions, for participants who were instructed to come up with practical collective action strategies to address violence against women, engaging in group discussion directly increased social identification as a supporter (see Figure 10 for direct path coefficients, values to the left of the forward slash). Identification negatively predicted defensiveness, which, in turn, negatively predicted collective action intentions. Tests of the indirect effect showed that there was a (marginally) significant indirect effect of engaging in group discussion on defensiveness via identification (IE = -.17, 95% CI [-.35, .01]). The expected indirect effect of group discussion on collective action intentions via identity *and* (latent) defensiveness was not significant (IE = 0.14, 95% CI [-.02, .30]).

I was puzzled about why the indirect effect was not significant when the pattern of coefficients clearly shows evidence consistent with mediation (see Figure 12). However, MacKinnon (2008; see also MacKinnon et al., 2010) suggests that this is not uncommon, and results from the fact that traditional symmetric confidence intervals produced by Mplus

cannot account for an indirect effect that is skewed. MacKinnon (2008, p. 99) therefore suggests that taking a bootstrapping approach can improve the accuracy of the confidence intervals for non-normally distributed indirect effects. Accordingly, I conducted an additional test of the indirect effect using 1,000 bias-corrected bootstrapped samples. When the indirect effect accounted for non-normality, the indirect effect of group discussion on defensiveness via identification became significant, (IE = -.17 95% CI [-.32, -.02]), as did the full indirect effect of group discussion on action via identity and defensiveness (IE = 0.14, 95% CI [.01, .28]). As Mplus does not allow for bootstrapping using multi-level models with nested data, it is important to note that tests of the indirect effect using bootstrapping were run while not accounting for the non-independence of the data (although the coefficients were virtually identical).

A rather different pattern of results was found for participants who were instructed to brainstorm about the *issue* of sexual harassment and assault against women (but not any strategies to address it). This model had adequate fit (SRMR = .07). Figure 12 shows that, for this condition, group interaction did not influence identification as a supporter (values to the right of the slash). Identification negatively predicted defensiveness, which, in turn, negatively predicted collective action intentions. For participants in this condition, there was no significant indirect effect of group discussion on collective action intentions via identification and defensiveness, (IE = 0.02, 95% CI [-0.03, 0.06]), or of group discussion on defensiveness via identification, (IE = -0.03, 95% CI [-0.13, 0.06]. This continued to be the case when the bias corrected bootstrapped confidence intervals were used to assess the indirect effect of group discussion on action via identification and defensiveness (IE = 0.02, 95% CI [-0.03, 0.07]); and on defensiveness via identification (IE = -0.04, 95% CI [-0.13, 0.05])
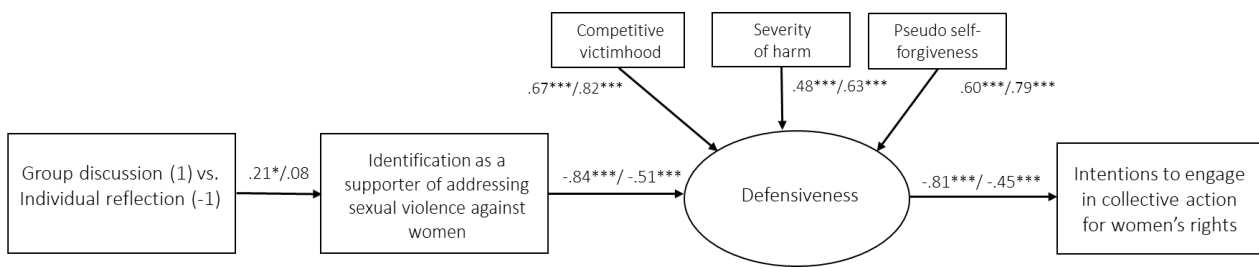
Figure 12. Results of the hypothesized path model with collective action intentions as the outcome, Study 8. Values to the left of the forward slash represent those in the action instructions condition, while values to the right represent those in the control instructions condition. Values are standardized regression coefficients (*p <.05, ***p < .001).

Behavioural measure of action. A similar pattern of results was obtained when I ran two mediation models using the behavioural measure of action as the outcome variable (whether or not participants clicked on a link relevant to expressing support for the Violence Against Women Act). For participants instructed to discuss practical collective action strategies to address the issue of gendered violence, engaging in group discussion boosted identification as a supporter of addressing sexual harassment and assault against women (values to the left of the forward slash, see Figure 13), SRMR = .04. Identification negatively predicted defensiveness (latent variable), which, in turn, was negatively associated with the behavioural measure of action. Initial tests of the indirect effects revealed that the indirect effect of group discussion on participants' behaviour via both identification and defensiveness was not significant, IE = .03, 95% CI [-.01, .08]. There was a significant indirect effect of group discussion on behaviour via identification, IE = -.12, 95% CI [-.24, .00]. However, when bootstrapped confidence intervals were used to model the indirect effect, the indirect effect of group discussion via both identification and defensiveness was significant, IE = .03, 95% CI [.00, .08].

Figure 13 (values to the right of the forward slash) shows that a different pattern of effects was obtained for participants instructed to reflect solely on the *issue* of sexual harassment against women, SRMR = .04. Here, engaging in group discussion had no significant impact on identification as a supporter of addressing sexual harassment and assault against women. Identification as a supporter negatively predicted defensiveness, which negatively predicted the behavioural outcome measure of action. There was no indirect effect of group discussion on defensiveness via identification as a supporter (IE = -.04, 95% CI [-.12, .05]), or on behavioural action via both identity and defensiveness (IE = .01, 95% CI [-.01, .03]). This continued to be the case when the bias corrected bootstrapped confidence intervals were used to assess the indirect effect: group discussion on defensiveness via identification as a supporter, IE = -.04, 95% CI [-.12, .05]; group discussion on behavioural action via both identity and defensiveness, IE = .03, 95% CI [-.01, .03].
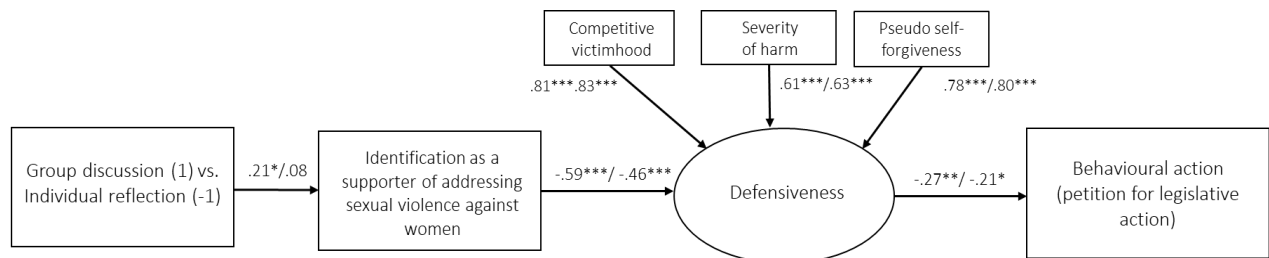

Figure 13. Results of the hypothesized path model with behavioural action as the outcome, Study 8. Values to the left of the forward slash represent those in the action instructions condition, while values to the right represent those in the control instructions condition. Values are standardized coefficients (*p* < .05, **p* < .01; ***p* < .001).

The pattern of results in Study 8 provides support for my hypothesized pattern of moderated mediation. Group discussion had positive indirect effects on collective action intentions *and* behaviour via both identity and latent defensiveness, but only for participants

who were instructed to reflect on practical solutions to address gender inequality. On the other hand, group discussion did not shape action (relative to individual reflection) when the task instructions focussed only on discussion of the issue. These results suggest that an increase in supporter identity and associated reductions in defensiveness are acting as key underlying mechanisms linking (mobilising) group discussions with men's commitment to act. Importantly, this was the case not only for men's *intentions* to act – the effects of group discussion also influenced men's behaviour to engage in action for women's rights. In line with the results of the LIWC analyses in Study 7, these results suggest that the effectiveness of group discussion to facilitate social identities for action is contingent on the extent to which participants' discussions are dedicated to discussing practical strategies for action. That is, it is not enough for participants to talk to one another about the importance of gender equality. Rather, interactions must involve discussion of men's plans to promote equality.

General Discussion

The current chapter reports on four studies investigating the impact of group discussion on men's identification as a supporter of combatting sexual violence, defensiveness, and commitment to act for gender equality. Overall, the results provide support for the idea that group discussion, and in particular, group discussion centred on concrete solutions to address violence against women, is a promising means of promoting men's commitment to progressive social change. Consistent with predictions, engaging men in group discussion regarding practical solutions to address gendered violence strengthened their social identification as a supporter of efforts to end sexual harassment and assault against women (Studies 5 and 8). Identification as a supporter was negatively associated with defensiveness (engagement in competitive victimhood, minimizing harmdoing, and pseudo self-forgiveness), and positively predicted men's collective action intentions and behaviour to address gender inequality (Studies 5, 6, and 8).

Importantly, these findings suggest that the *content* of men's interactions was key in influencing their commitment to act for women's rights. In Study 6, group discussion had little impact on men's opinion-based identification as a supporter of efforts to address violence against women, defensiveness, or collective action. Analysis of the explicit content of the interactions in Study 5 and 6 revealed that participants' discussions differed regarding the prevalence of collective action relevant language expressed. That is, on average, participants in the discussion groups in Study 5 talked more about collective action strategies than participants who engaged in group discussion in Study 6. Study 8 therefore experimentally tested this boundary condition explicitly to reveal that group discussion (relative to individual reflection) boosted identification *only* for participants who were instructed to brainstorm concrete strategies to address sexual violence. On the other hand, group discussion had little effect on identification as a supporter for participants who were instructed to reflect solely on the issue of sexual harassment (i.e., why the issue is important).

Taken together, the current chapter provides initial evidence regarding the fruitfulness of taking a group interaction approach to (a) reduce men's defensiveness regarding the issue of sexual harassment and violence against women and (b) increase their collective action intentions (and behaviour) for gender equality. Importantly, Studies 5 and 8 provide support for both social identity *and* defensiveness as key mechanisms underlying the relationship between group interaction and men's commitment to action. That is, through talking with other men about addressing the issue of violence against women, men came to see themselves and others as belonging to the same social group defined by a desire to challenge gender inequality. As a result, their defensiveness regarding the issue of gendered violence was lowered, which, in turn, was associated with their increased collective action for gender equality.

The current findings align with previous research showing that group interaction is a facilitator of meaningful social identities grounded in opinions about social and political issues (Postmes et al., 2005; Thomas & McGarty, 2009; Thomas et al., 2016). The current study, however, goes beyond existing findings by being the first (to my knowledge) to explore the effects of group interaction as a means of mitigating men's defensiveness regarding the issue of gendered violence (and, consequently, engendering their support for collective action for gender equality). The findings showed that group interaction not only (indirectly) influenced men's intentions to engage in collective action, but that it also positively predicted observed behavioural change (men's engagement in action related to bringing about legislative change to protect survivors of domestic and sexual violence). This finding is particularly promising, given that people's intentions are sometimes an imperfect predictor of their actual behaviour (Sheeran & Webb, 2016).

Significantly, the findings of the current studies show that even relatively brief group discussions (15 minutes) conducted online were enough to lead to psychological change (social identity formation), which, in turn, positively predicted men's collective action intentions and behaviour. The current results thereby extend research that has previously only used face-to-face discussion paradigms by demonstrating the success of targeted group discussion online, where participants were all anonymous and unknown to one another. Indeed, research conducted within the framework of the Social Identity Model of Deindividuation Effects (SIDE; Reicher et al., 1995) suggests that processes of social influence can be even stronger in anonymous online environments than during face-to-face interaction. SIDE proposes that in the absence of individuating information (e.g., usernames, profile pictures) people become increasingly sensitive to the norms operating within a given context (Postmes & Spears, 1998; Reicher et al., 1995; Spears et al., 1990). However, whether online group discussion interventions (via synchronous chat or video

teleconferencing apps such as Skype or Zoom) may be *more* successful in targeting behaviour than face-to-face interventions remains a potential avenue for future research.

The current findings have important practical implications for interventions seeking to involve men in the fight for gender equality – as well as for strategies that seek to reduce prejudice against women and disadvantaged groups more generally. Specifically, the results suggest that diversity, equity, and inclusion (DEI) initiatives that aim to address sexual harassment in the workplace are more likely to result in concrete behavioural change to the extent that men are able to come together to discuss practical solutions to address the problem. That is, it is not enough to get people to individually reflect on the issue at hand, or even discuss why the issue is important (in supportive terms) with others. Rather, men need to be engaged in interactions where they share concrete plans for action in order for a sense of common identity to emerge. Of course, the degree to which this identity is predicated on prosocial norms relies heavily on the content of such group interactions which – in the context of implementing successful group discussion interventions within the workplace – would need to be closely guided and monitored.

Limitations and Future Directions

As in Chapter 3, many of the effects reported in the current chapter are correlational. While group discussion had direct effects on opinion-based identification as a supporter (Study 5 and Study 8), pseudo self-forgiveness (Study 5), collective action intentions (Study 5), and collective behaviour (Study 5), it had no direct effects on defensive outcomes or collective action outcomes consistently across studies. Rather, the effects on defensiveness and action occurred indirectly via men's social identification as a supporter. As a result, future research should focus on replicating the results obtained here to ensure the reliability – as well as the generalisability – of the current findings. For example, future research could

test the efficacy of group discussion interventions in other intergroup contexts (e.g., the movement to end racial inequality and injustice).

Further, it is unclear whether group discussion has any long-term effects on men's identity, defensiveness, and action intentions. Research in the context of the Syrian refugee crisis demonstrated that when online discussions regarding harm against refugees were ongoing, people remained committed to expressing solidarity and promoting justice for refugees (Smith et al., 2018). Future research could therefore examine whether the effects found in the current studies (Study 5 and Study 8) extend to men's continued commitment to act for gender equality in the future.

Finally, although the current findings go some way in addressing the processes underlying the relationship between group discussion and collective behaviour, there are other aspects of group interaction that I did not explicitly examine and that warrant further investigation. In particular, previous research has emphasized the importance of group members' consensus and validation as key qualifying factors of the identity formation process (Thomas et al., 2019; Smith & Postmes, 2011). Although examining consensus and social validation was outside the scope of the current chapter, future research should attend to the influence of both opinion-consensus and opinion-validation to provide a more comprehensive understanding of when group discussions targeting prosocial behaviour will be successful. Importantly, it is likely that the processes described here in the context of men's *prosocial* action may also underly their participation in more hostile forms of anti-feminist action. Thus, future research could also seek to examine the group dynamics that facilitate men's engagement in *reactionary* forms of action (e.g., identification as a Men's Rights Activist; Schmitz et al., 2016).

Conclusion

In this chapter I highlight how group interaction interventions could be well positioned to tackle advantaged group members' defensiveness and increase their intentions to engage in social movements for justice and equality. Importantly, the conceptual and empirical evidence presented here encourages the use of social solutions to tackle inherently social problems (see also Fielding & Hornsey, 2016; Thomas et al., 2016; Prentice & Paluck, 2020; Ryan, 2022; Smith et al., 2020a). Men's views regarding modern feminist movements do not arise in a vacuum – rather, they are constantly influenced by discussions within their immediate social networks, which are, in turn, shaped by larger social structures that embed and reproduce inequality.

Chapter 5

General Discussion

The broad aim of this thesis was to explore the social psychological factors that underlie men's support for (versus their resistance to) collective action for gender equality. I have used this question to explore practical issues related to promoting men's support for gender equality, but also theoretical and empirical issues regarding the dynamic and interdependent processes that influence men's support for progressive and reactionary forms of social change. Specifically, I have provided evidence that online opposition to discussions about violence against women are often characterized by attempts to defend men's moral image. I showed that men's concerns about their group's morality and their defensive construals regarding gender-based violence can undermine their intentions to act for women's rights – as well as motivate their intentions to advocate for the rights of their own group. Importantly, I demonstrated that engaging men in group discussion regarding strategies to address gender inequality can provide a means of overcoming defensive reactions and boosting their commitment to act to address violence against women.

The present thesis has implications for both theory and practice regarding the factors that may promote (or hinder) men's commitment to act for gender equality. The findings suggest that men's distinct identity-based needs may contribute to their countermobilization to oppose feminist efforts. However, the results also offer a fruitful approach to mitigate defensive reactions and engage men in efforts to address gender inequality. In the present chapter I discuss the implications of these findings for our understanding of both theory and practice related to fostering advantaged group members support for equality. Finally, I consider the limitations of this thesis, and suggest some directions for future research to further our understanding of advantaged group members' support for progressive social change.

Defensiveness influences men's intentions to challenge and support the status quo

A key contribution of the present thesis was a focus on examining men's support and resistance to gender equality as two sides of the same coin – that is, I asked whether the factors that may reduce men's intentions to act for gender equality (i.e., their support for progressive or *system-challenging* forms of action) may also motivate their participation in reactionary (or *system-supporting*) action to oppose social change. In Chapter 2, I provided evidence that online opponents of the movement to end violence against women commonly expressed concerns about the impact of sexual assault allegations on men's moral reputation. Using a topic modelling approach, I identified that, in the context of discussions about the issue of men's violence against women, Twitter users employed a range of strategies to defend men's moral reputation – including victim-blaming, competing with women for victim status, and minimizing the structural nature of gender-based violence by claiming that it is only perpetrated by a deviant few members of the ingroup.

The results of Chapter 2 extend previous literature on defensive reactions to reminders of ingroup harm by providing evidence of defensive rhetoric being mobilized in naturally occurring social media data. Previous research has tended to look at the use of defensive strategies within the context of groups involved in intractable violent conflict and/or historical memory for wrongdoing (Bilali et al., 2012; Noor et al., 2008; 2012; Peetz et al., 2010; Rotella & Richeson, 2013; Shnabel et al., 2013a; Uluğ et al., 2021). Much of this work has relied on experimental/cross-sectional methods using participant's self-reported responses to scenarios, which may underestimate the occurrence of advantaged group members' defensiveness (due to social desirability concerns), as well as limit our understanding of how defensive rhetoric is strategically employed in the real world. Further, the collective action literature has considered how challenges to the status quo can encompass a variety of different actions, captured broadly in the distinction between *normative* and *nonnormative*

collective action, respectively (Tausch et al., 2011; Teixeira et al., 2020). However, less work has sought to distinguish between the various actions people can take to *uphold* the prevailing social system (as a form of *system-supporting* or *reactionary action*; Jost et al., 2017). By explicitly examining the tactics used by those who seek to push back against the movement for gender equality, the findings of the present thesis highlight how action to oppose feminist efforts can include a range of defensive claims regarding men's moral character.

Given the centrality of attempts to defend men's moral image in tweets from opponents in Chapter 2, in Chapter 3 I explicitly tested whether men's intentions to support or oppose action for women's rights is shaped by their concerns about their group's morality (and their subsequent need to defend their ingroup). Studies 2, 3, and 4 showed that men's social identification was associated with their heightened need for morality (in line with the tenets of the needs-based model; Shnabel & Nadler, 2008). Need for morality, in turn, was positively associated with defensive construals regarding the issue of sexual harassment and assault (captured by competitive victimhood, minimization of harm, and pseudo self-forgiveness). Defensiveness was *negatively* associated with men's intentions to engage in action to address sexual harassment against women and was *positively* associated with men's intentions to promote the rights of their own group. Importantly, defensiveness stemmed from men's commitment to their social group – although there was no evidence that this relationship was heightened by exposure to women's explicit accusations.

The current findings bridge together the literature on defensive responses to reminders of ingroup harm with the literature on collective action to suggest that men's need to defend their group's morality may not only reduce their intentions to act for women's rights – but also contribute to men's countermobilization to promote the rights of their own group. These findings seem particularly significant, given that supporters of reactionary counter-movements to feminist efforts (such as "#MenToo" or "#HimToo") have argued that the use

of these hashtags simply represent attempts to broaden the inclusivity of the gender equality movement (Boyle & Rathnayake, 2020). However, the present findings show that men's intentions to act on behalf of their own group were associated with their attempt to downplay the issue of violence against women, blame women for the issue of sexual harassment and assault, and claim that men now suffer *more* in society (relative to women). These results align with recent work conducted by West et al. (2021) in the context of racial inequality – who showed that while supporters of "All Lives Matter" (ALM) have similarly argued that ALM is more inclusive than "Black Lives Matter", support for ALM was driven by colour-blind ideologies that seek to deny the reality of racial inequality.

Taken together, the results emphasize the need for research on collective action and intergroup relations to consider how advantaged group members' concerns about their ingroup's morality may facilitate their engagement in reactionary forms of action. To date, research exploring reactionary or system-supporting action has tended to focus on how threats to advantaged group members' power and privilege can motivate their opposition to movements for social change (Hodson et al., 2022). However, the results of Chapter 2 and 3 suggest that men's morality needs (and subsequent defensiveness) may uniquely contribute to their countermobilization. Indeed, Okuyan & Vollhardt (2022) note that people's resistance to social change need not be overtly violent to cause harm. That is, while defensive reactions may *appear* to be less harmful than more violent forms of intergroup resistance, they are dangerous precisely because of how they subtly work to obscure the reality of group-based inequalities, and as a result, cast doubt on the necessity of social change.

The power of group processes in overcoming defensiveness and promoting men's action for gender equality

The second key contribution of this thesis was demonstrating how social identity processes can be harnessed to overcome defensive reactions and promote men's commitment

to act for gender equality. In Chapter 4, engaging men in group discussion regarding strategies to combat gender inequality boosted their identification with a group grounded in support for addressing the issue of gender-based violence. Identification as a supporter was negatively associated with defensive construals regarding the issue of gender-based violence, which, in turn, had positive flow on effects for men's action.

Importantly, the current findings emphasize that the *content* of people's interactions is critical in shaping identities oriented towards prosocial action (Thomas et al., 2022a; Smith et al., 2020b). The studies in Chapter 4 showed that it was not enough for men to discuss *why* the issue of sexual violence against women is important. Rather, it was only when men discussed *how* to address the problem of sexual violence – through sharing concrete actions they could take in their everyday lives – that group discussion influenced identification, defensiveness, and collective action. Mobilising group discussions were not only positively associated with men's intentions to act for gender equality, but also directly influenced participant's behavioural change (they were more likely to click on a link to a webpage relevant to expressing support for the Violence Against Women Act). The positive effects of group discussion on the behavioural measure of action in Study 5 and 8 add significant weight to the current findings, given people's behaviour is a much stronger indicator of the efficacy of a particular intervention than mere intentions alone (Sheeran & Webb, 2016).

The findings of Chapter 4 contribute to the existing literature on the transformative power of group discussion on people's commitment for collective action (Postmes et al., 2005; Thomas et al., 2016; 2019; Thomas & McGarty, 2009; Smith & Postmes, 2009; 2011; Smith et al., 2015; 2020b). However, the current studies go beyond previous research by demonstrating that the effects of group discussion can be extended to the domain of gender inequalities between men and women. Further, the results suggest that group discussion (and

its effects on the formation of social identities crafted for action) can be strategically deployed to target the problem of men's defensiveness regarding gender-based violence.

The findings contribute to the broader social psychological literature by emphasizing the importance of accounting for group processes when designing solutions to promote widespread behavioural change (Fielding & Hornsey, 2016; Thomas et al., 2016; Prentice & Paluck, 2020; Ryan, 2022; Smith et al., 2020a). Historically, social psychology as a discipline has tended to focus on exploring social *problems* (for example, intergroup prejudice, hostility, violence), with less work concentrated on developing *solutions* to these problems. Although research has made strides in designing psychological interventions to promote intergroup reconciliation, these studies have largely been conducted in the context of intractable violent conflicts between groups (e.g., Shnabel et al., 2013a). Further, psychological studies targeting outgroup prejudice have tended to individualise the issue of prejudice, that is, prejudice is viewed as something that resides *within* a person, rather than constructed, embedded, and reproduced socially (Durrheim et al., 2016; Perez & Salter, 2020; Platow et al., 2019). The present thesis therefore extends previous work by offering an approach to addressing inequality that is grounded firmly within a social identity account of group processes, which considers the way that men's identities, attitudes, and behaviours are dynamically shaped by their micro- and macro- level interactions.

Although my focus here has been to discuss the findings of Chapter 4 in the context of men's support for progressive social change, it is important to acknowledge that they may also explain the processes through which men come to participate in more hostile forms of anti-feminist action. Research has pointed to social media platforms as an importance space for misogynistic subcultures to take root (Rafail & Freitas, 2019; see also Chapter 2). However, less research has examined the intragroup processes (i.e., social identity formation) that facilitate men's involvement in these communities. Future research should therefore seek

to explore how group processes contribute to men's endorsement of extreme misogynistic ideologies (for example, men's identification within the "incel" or *involuntary celibate* community). This investigation seems particularly pertinent, given growing concerns regarding the influence of these online communities on men's decision to engage in violence against women offline (Helm et al., 2022; Srinivasan, 2021; Witt, 2020).

Triangulating different methodologies in the study of support for and action against social change

A significant strength of the present thesis was the use of multiple (pre-registered) methodologies to triangulate how both inter- and intragroup processes may shape men's support for (and resistance to) gender equality. I used a combination of experiments (including a novel online group-discussion paradigm), computational analyses (topic modelling Twitter data), and natural language processing to capture the complexity and dynamism of men's support for progressive (versus reactionary) social change efforts. The strengths and limitations of each of these methods offset the other. That is, while social media data has high ecological validity, it only captures a particular sub-set of interactions among already politicized people (and is therefore not wholly representative of the general population). Conversely, online experiments allow for the isolation of psychological phenomena, thereby resulting in stronger internal validity (but weaker external validity). However, as I note below, experimental paradigms that include *actual* interaction between participants allows for the systematic study of group processes while not compromising on ecological validity.

In Chapter 2, I used a computational approach to analyse the nature of online discussions regarding the issue of gender-based violence. By comparing the tactics used in tweets from both supporters and opponents, I highlighted how social networking sites are a key site of modern intergroup contact (and conflict) between groups. Historically, the

literature on intergroup contact has operationalized contact primarily as something that occurs *offline* (e.g., in the workplace, in people's interpersonal relationships; Kende et al., 2018; Zhou et al., 2018; but see White et al., 2019; 2020; 2021; for some exceptions). However, I reason that this is a narrow conceptualization of contact that neglects to account for how social media connects people from diverse social and ideological groups (Castells, 2012; Lüders et al., 2022; Thomas et al., 2022a). Indeed, the ideological diversity of social media has been implicated as a key factor driving increasing online political polarization (Bail et al., 2018). Thus, I would urge future research to explore the nature of *online* contact between groups – whether characterized by pre-existing memberships (e.g., gender or racial identity) – or groups grounded in ideologies or opinions about the world (e.g., identification as a feminist). I argue that methodologies designed to capture both *where* and *how* people are interacting about social and contentious political issues is critical to developing a comprehensive understanding of the dynamic and interdependent nature of people's support and resistance to social change. The accessibility of open-source data (such as the tweets collected using Twitter's API for academic research in Chapter 2) allows for the analysis of social interactions in ways that have been previously unavailable to researchers.

Chapter 3 used a series of experimental studies designed to specifically isolate how the nature of women's claims (the extent to which women highlighted men's responsibility for perpetrating sexual harassment) influenced men's support and resistance to gender equality. The fact that the three empirical studies in Chapter 3 failed to successfully manipulate a threat to men's need for morality highlights a key challenge of experimental social and political psychology: that it is often difficult to shift the nature of people's deeply rooted identities, ideologies, attitudes, and behaviours. As outlined in Chapter 3, the failed manipulations may have been due to methodological issues within those studies, but also may be because the issue of sexual harassment is chronically threatening to men's moral identity.

However, these studies were important in providing an empirical test of some of the findings that emerged from the topic modelling analysis used in Chapter 2.

In the final empirical chapter of this thesis, I used a group discussion paradigm – whereby participants were involved in online synchronous chat with one another – to test the role of intragroup interaction on men's social identification as a supporter to address women's rights, defensiveness, and collective action intentions and behaviour. I complemented this approach by using natural language processing to explore whether the specific content of participants' interactions was related to their action for women's rights. Research in social psychology has tended to shy away from using social interaction paradigms, despite the significant benefits such methodologies offer for capturing the interactive and dynamic nature of people's identities, attitudes, and behaviour (Haslam & McGarty, 2001). The online and unmoderated nature of the group chat program in the studies presented within this thesis also increase the ecological validity of the findings – which may have otherwise been compromised in designs that use confederates or where interaction occurs within a laboratory environment. As a result, the current findings are particularly well positioned to inform policy and practice related to motivating men's commitment to positive social action. It is to this point that I now turn.

Implications for policy and practice

The findings of this thesis provide implications for policy and practice regarding how to increase men's solidarity-based action for gender equality. First and foremost, the results emphasize the need for practitioners and policy-makers to account for the power of group processes when developing programs aimed at engaging men in action for progressive social change (Bavel et al., 2020; Drury & Reicher, 1999; 2000; 2009; Haslam et al., 2020; Thomas et al., 2022a; Reicher et al., 2020). While the findings suggest that men's identification with their broad gender category may lead to resistance (due to concerns about their group's moral

reputation; Chapter 3), framing men to be a part of the *solution* offers a promising avenue to overcome defensive reactions and mobilise men's commitment to act. The findings of Chapter 4 suggest that this mobilization can occur when men get together to talk to one another (either face to face, or online) about concrete strategies for action. Through sharing ideas about how to combat sexual harassment and assault, men can come to define themselves in terms of an identity based in support for the movement to address gender-based violence. In line with the insights of the solidarity-based model of social change (Subašić et al., 2008; 2018), it is through this process that men may come to see themselves as "agents of change" – rather than as bystanders or perpetrators (Subašić et al., 2018, p. 709).

In addition to examining these micro-level dynamics (that is, men's small group discussions with one another), it is important to consider how the content of men's intragroup discussions are shaped by broader societal discussions regarding gender equality. The existing literature on identity-formation and leadership suggests that it is not just the *content* of people's interactions with one another that is important (i.e., what is being said), but also, *who* is saying it (Haslam et al., 2017; 2020; Reicher et al., 2005; Subašić et al.; 2018, p. 709). The work of Reicher et al. (2005) highlights how public figures are central in shaping people's commitment to social change by signalling the appropriate and desired behaviour within a given context. In the context of men's participation in action for gender equality, Subašić et al. (2018) propose that powerful men shape other men's commitment to act by communicating a *shared* identity that includes both men and women as supporters of efforts to address gender inequality. Taken together, I propose that initiatives seeking to mobilise men for social change should attend closely to the mutually constitutive relationship between men's private and public discussions regarding gender inequality. That is, the current findings suggest that it is not enough to ask men working in a particular organisation to discuss strategies to address gender inequality if they are not receiving similar messages from their

leaders and broader institution. Rather, widespread positive behavioural change is most likely to occur when our leaders and institutions are involved in cultivating a shared sense of identity (a sense of "us-ness"; Haslam et al., 2020) that can be used to actively reshape norms, attitudes, and behaviour towards challenging discrimination and injustice.

Limitations to the present research program and directions for future research

The current thesis has reinforced the idea (well known to researchers studying human behaviour) that understanding people's commitment to social change is dynamic and complex (Dixon et al., 2020; Drury & Reicher, 1999; Stott & Drury, 1999; Reicher, 1996). In order to further our understanding of advantaged group members' support and resistance to movements advocating for social progress, it is important to continue developing methodologies that can better capture the complexity of the phenomena we are interested in. Thus, in the following section I note some limitations of the current thesis and propose directions for future research to meet this aim.

At the beginning of this thesis, I argued for collective action to be understood as a dynamic, intergroup phenomenon (as per Drury & Reicher, 1999; 2000; Postmes et al., 2005; Reicher, 1996; 2004). While the work in this thesis has gone some way in achieving this end, there are important future avenues for research to capture the interactive and dynamic nature of collective action for and against social change. For example, in Chapter 2 I explored the contentious nature of online interactions regarding sexual harassment and assault against women. While the topic modelling approach used in this chapter allowed me to identify key themes within the corpus of tweets, this analysis did not account for how movements and counter-movements on social media evolve dynamically over time. Future research could therefore use an approach that can capture and map people's engagement over the course of a movement – and/or include an exploration of how people's offline protest influences their online interactions, and vice versa (Thomas et al., 2022b). For example, a network analysis

approach could be used to chart peoples' activity within a particular social networking platform over time – as well as the nature of this activity (e.g., the use of popular hashtags) and users' connection to (and influence over) other users.

Within the scope of this thesis I explored how men's responses to gender inequality may be shaped by intergroup interactions (women's claims, morality needs, and defensiveness; Chapter 3) and intragroup interactions (men's discussions with one another; Chapter 4). However, an important avenue for future research would be to consider the perspective of women, alongside men. That is, in order to truly account for the intergroup nature of collective action – research should consider the perspectives of both disadvantaged and advantaged group members, side by side. In the context of encountering men's backlash to allegations of violence against women, it may be that women experience a further threat to their power, leading them to take more "vengeful" tactics in response as a means of gaining control (Strelan et al., 2014). Conversely, opposition from men may undermine women's perceptions of group efficacy (that is, their belief that coordinated group action can be effective in creating change), thus leading them to disengage from collective action efforts (in line with the social identity model of collective action; SIMCA; van Zomeren et al., 2008). The use of methodologies to examine the interplay between advantaged and disadvantaged group members' claims and established predictors of collective action (anger, identity, group efficacy) – as well as group member's identity needs (power, morality) – provides an interesting avenue for future research. Such studies could include designs that allow for *intergroup* interaction between group members, in order to mirror how these online discussions evolve dynamically and interdependently over time.

Of course, as I have raised earlier within this thesis, support for social change is not clearly divided across the boundaries of "advantaged" or "disadvantaged" group membership (Dixon et al., 2020; Thomas et al., 2022a; Siem et al., 2016). Advantaged group members

often engage in action to support disadvantaged group members, and disadvantaged group members can engage in actions that seemingly go against their group interests (e.g., anti-feminist women; Mikołajczak et al., 2022). In Chapter 4 I outlined how this action is best explained by people's identification with groups based in shared ideologies or *opinions* about the state of the world – that cut across pre-defined social categories (e.g., race or gender; Bliuc et al., 2007; McGarty et al., 2019; Smith et al., 2015; Subašić et al., 2008; Thomas et al, 2009; 2022a). A nuanced account of support for both progressive versus reactionary collective action should therefore consider social identities that transcend traditional intergroup boundaries, as well as how people's membership in multiple groups can intersect to shape their support for and against social change efforts.

Indeed, it is important to note that the experimental studies within this thesis (Chapters 3 and 4) took a single-identity approach to understanding advantaged group identity by solely recruiting participants who identified as men. In this way, the studies did not account for how other group memberships – such as race or sexual orientation – intersect with gender identity to influence people's responses to inequality (Bowleg et al., 2017; Howard & Renfrow, 2014). In the case of the #MeToo movement, sexual assault allegations predominantly focused on cis-gendered, White, heterosexual women as victims, and cis-gendered White, heterosexual men as the perpetrators (Leung & Williams, 2022). Further, there is a history of white women falsely accusing Black men of rape (Srinivasan, 2021). It is therefore likely that perceptions of sexual violence against women differ depending on people's identification in multiple groups, and the way that these identities intersect to further advantage (or disadvantage) individuals. Although participants in the experimental studies in Chapters 3 and 4 predominantly identified as cis-gendered, White, and heterosexual, participant recruitment was not explicitly limited to men who identified within these categories. Thus, while the results of Chapter 3 link men's gender identification to their

support and resistance, it is likely that this relationship is influenced by men's intersecting group memberships. Future research should therefore consider the intersectional nature of privilege to provide a more nuanced understanding of responses to debates surrounding gender inequality (see also Cole et al., 2009; Nair & Vollhardt, 2020; Settles & Buchanan 2014).

A final point concerns the generalizability of the current findings to other contexts of structural inequality between groups. While my focus in the present thesis has been to explore how men (as members of a structurally advantaged group) respond to efforts to progress the movement for gender equality, it is important that models of collective action explain behaviour across a variety of intergroup contexts (including those in non-Western/non-WEIRD countries; Henrich et al., 2010). Thus, future work should examine whether the key findings of this thesis are applicable to other intergroup contexts. For example, future studies could see whether intragroup interaction among White participants regarding strategies to address racial inequality could effectively mobilise their engagement in collective action to address injustice against racial minorities. Future research could extend the results found here across a variety of intergroup contexts characterized by inequality, as well as exploring unique aspects of these contexts that may act as boundary conditions to the processes outlined within this thesis.

Conclusion

At a time of renewed attacks on reproductive freedoms, LGBTIQA+ rights, and ethnic minority group members, engaging advantaged group members in the fight for justice for disadvantaged groups seems particularly pressing. This thesis has put forward some potential answers for how we may overcome advantaged group members' resistance to social change, and instead, promote their commitment to act for equality. As bell hooks (1995) writes, "There must exist a paradigm, a practical model for social change that includes an

understanding of the ways to transform consciousness that are linked to efforts to transform

structures" (p. 118). It is my hope that the findings of this thesis have contributed to this

challenge.

References

Alberici, A. & Milesi, P. (2016). Online discussion, politicized identity, and collective action. *Group Processes & Intergroup Relations, 19*, 43-59. https://doi.org/10.1177/1368430215581430

Alcoff, L. M. (2018). *Rape and resistance*. London: Polity Press.

Ali, S. H., Mohaimin, S., Dhar, R., Dhar, M., Rahman, F., Roychowdhury, L., ... & Lim, S. (2022). Sexual violence among LGB+ South Asian Americans: Findings from a community survey. *Plos One*, *17*, 1-7. https://doi.org/10.1371/journal.pone.0264061

Arun, R., Suresh, V., Veni Madhavan, C. E., & Murthy, N. (2010). On finding the natural number of topics with latent dirichlet allocation: Some observations. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 391–402). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13657-3_43

Aydin, A. L., Ullrich, J., Siem, B., Locke, K. D., & Shnabel, N. (2019). Agentic and communal interaction goals in conflictual intergroup relations. *Journal of Social and Political Psychology*, *7*, 144–171. https://doi.org/10.5167/uzh-170746

Baele, S. J., Brace, L., & Coan, T. G. (2021). From "Incel" to "Saint": Analyzing the violent worldview behind the 2018 Toronto attack. *Terrorism and Political Violence*, *33*, 1667–1691. https://doi.org/10.1080/09546553.2019.1638256

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., ... & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*, 9216-9221. https://doi.org/10.1073/pnas.1804840115

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*, 193–209. https://doi.org/10.1207/s15327957pspr0303_3

Bandura, A. (2011). Moral disengagement. *The encyclopedia of peace psychology*. Wiley.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral
disengagement in the exercise of moral agency. *Journal of Personality and Social
Psychology*, *71*, 364–374. https://doi.org/10.1037/0022-3514.71.2.364

Barron, A. T., & Bollen, J. (2021). Quantifying Collective Identity Online from Self-defining
Hashtags. Unpublished manuscript. https://doi.org/10.21203/rs.3.rs-960863/v.

Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., ... & Willer,
R. (2020). Using social and behavioural science to support COVID-19 pandemic
response. *Nature Human Behaviour*, *4*, 460-471. https://doi.org/10.1038/s41562-020-
0884-z.

Becker, J. C. (2020). Ideology and the promotion of social change. *Current Opinion in
Behavioral Sciences*, *34*, 6–11. https://doi.org/10.1016/j.cobeha.2019.10.005

Belknap, J. (2010). Rape: Too hard to report and too easy to discredit victims. *Violence
Against Women, 16,* 1335–1344. https://doi.org/10.1177/1077801210387749

Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and
Individual Differences*, *42*, 825–829. https://doi.org/10.1016/j.paid.2006.09.024

Bhattacharyya, R. (2018). #Metoo movement: An awareness campaign. *International Journal
of Innovation, Creativity and Change*, *3*, 1–12. https://doi.org/10.1111/aswp.12099

Bilali, R. (2013). National narrative and social psychological influences in Turks' denial of
the mass killings of Armenians as genocide. *Journal of Social Issues*, *69*, 16–33.
https://doi.org/10.1111/josi.12001

Bilali, R. (2022). Fighting violent extremism with narrative intervention: Evidence from a
field experiment in West Africa. *Psychological Science, 33,* 184-195.
https://doi.org/10.1177/09567976211031895

Bilali, R., Tropp, L. R., & Dasgupta, N. (2012). Attributions of responsibility and perceived harm in the aftermath of mass violence. *Peace and Conflict: Journal of Peace Psychology*, *18*, 21–39. https://doi.org/10.1037/a0026671

Bilali, R. & Vollhardt, J. R. (2019). Victim and perpetrator groups' divergent perspectives on collective violence: Implications for intergroup relations. *Political Psychology, 40,* 75–108. https://doi.org/10.1111/pops.12570

Blais, M., & Dupuis-Déri, F. (2012). Masculinism and the antifeminist countermovement. *Social Movement Studies*, *11*, 21–39. https://doi.org/10.1080/14742837.2012.640532

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bliuc, A. M., McGarty, C., Reynolds, K., & Muntele, D. (2007). Opinion-based group membership as a predictor of commitment to political action. *European Journal of Social Psychology, 37,* 19–32. https://doi.org/10.1002/ejsp.334

Bliuc, A. M., Smith, L. G., & Moynihan, T. (2020). "You wouldn't celebrate September 11": Testing online polarisation between opposing ideological camps on YouTube. *Group Processes & Intergroup Relations*, *23*, 827-844. https://doi.org/10.1177/1368430220942567

Bongiorno, R., McGarty, C., Kurz, T., Haslam, S. A., & Sibley, C. G. (2016). Mobilizing cause supporters through group-based interaction. *Journal of Applied Social Psychology*, *46*, 203–215. https://doi.org/10.1111/jasp.12337

Bovaird, J. A. (2007). Multilevel structural equation models for contextual factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 149–182). Lawrence Erlbaum Associates Publishers.

Bowleg, L. (2017). Intersectionality: An underutilized but essential theoretical framework for social psychology. In the Palgrave *Handbook of critical social psychology* (pp. 507-529). Palgrave Macmillan, London.

Boyle, K., & Rathnayake, C. (2020). #HimToo and the networking of misogyny in the age of #MeToo. *Feminist Media Studies*, *20*, 1259–1277. https://doi.org/10.1080/14680777.2019.1661868

Branscombe, N. R., & Miron, A. M. (2004). Interpreting the ingroup's negative actions toward another group: Emotional reactions to appraised harm. In L. Z. Tiedens & C. W. Leach (Eds.), *Studies in emotion and social interaction. The social life of emotions* (pp. 314-335). New York, NY, US: Cambridge University Press.

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.

Brooks, A. W. (2015). *ChatPlat: A New Tool to Study Real Human Interaction*. https://spsp.org/news-center/character-context-blog/chatplat-new-tool-study-real-human-interaction.

Burt, M. R. (1980). Cultural myths and support for rape. *Journal of Personality and Social Psychology, 38,* 217–230. https://doi.org/10.1177/1077801210387749

Burley, S., & Ross, A. R. (2019). From nativism to White power: Mid-twentieth-century White supremacist movements in Oregon. *Oregon Historical Quarterly, 120,* 564-587. https://doi.org/10.1353/ohq.2019.0030.

Byrne, B. M. (2013). *Structural equation modelling with EQS: Basic concepts, applications, and programming*. Routledge.

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, *72*, 1775–1781. https://doi.org/10.1016/j.neucom.2008.06.011

Castells, M. (2012). *Networks of outrage and hope: Social movements in the age of the Internet*. Polity Press.

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*, 899-920. https://doi.org/10.1111/josi.12408.

Čehajić-Clancy, S., Effron, D. A., Halperin, E., Liberman, V., & Ross, L. D. (2011). Affirmation, acknowledgment of in-group responsibility, group-based guilt, and support for reparative measures. *Journal of Personality and Social Psychology*, *101*, 256–270. https://doi.org/10.1037/a0023936

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems*, *22*, 1–9.

Charney, M. W. (2021). Myanmar coup: how the military has held onto power for 60 years. *The Conversation*. https://theconversation.com/myanmar-coup-how-the-military-has-held-onto-power-for-60-years-154526.

Choma, B., Hodson, G., Jagayat, A., & Hoffarth, M. R. (2020). Right-wing ideology as a predictor of collective action: A test across four political issue domains. *Political Psychology*, *41*, 303–322. https://doi.org/10.1111/pops.12615

Clark, M. D. (2019). White folks' work: digital allyship praxis in the #BlackLivesMatter movement. *Social Movement Studies, 18*, 519-534. https://doi.org/10.1080/14742837.2019.1603104.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences (2nd ed., pp. 109–143). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cole, E. R. (2009). Intersectionality and research in psychology. *American Psychologist, 64,* 170-180. https://doi.org/10.1037/a0014564.

De Maricourt, C. D., & Burrell, S. R. (2021). #MeToo or #MenToo? Expressions of backlash and masculinity politics in the #MeToo era. *The Journal of Men's Studies*, *30,* 49–69. https://doi.org/10.1177/10608265211035794

Demirdağ, A., & Hasta, D. (2019). Threat perception and perspective taking as mediators between competitive victimhood and evaluations of collective action: The Gezi Park protests. *Political Psychology*, *40*, 953–971. https://doi.org/10.1111/pops.12564

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Numérique*, *17*, 61–84. https://doi.org/10.3166/dn.17.1.61-84

Dickel, V., & Evolvi, G. (2022). "Victims of feminism": exploring networked misogyny and #MeToo in the manosphere. *Feminist Media Studies*, 1–17. https://doi.org/10.1080/14680777.2022.2029925

Dixon, J., Elcheroth, G., Kerr, P., Drury, J., Al Bzour, M., Subašić, E., & Green, E. G. (2020). It's not just "us" versus "them": Moving beyond binary perspectives on intergroup processes. *European Review of Social Psychology, 31*, 40-75. https://doi.org/10.1080/10463283.2020.1738767.

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

Douglas, M. (2021). Defamation actions and Australian politics. In *University of New South Wales Law Journal Forum* (pp. 1–12). Sydney, NSW: University of New South Wales, Faculty of Law. https://doi.org/10.2139/ssrn.3857025

Dover, T. L., Major, B., & Kaiser, C. R. (2016). Members of high-status groups are threatened by pro-diversity organizational messages. *Journal of Experimental Social Psychology*, *62*, 58–67. https://doi.org/10.1016/j.jesp.2015.10.006

Droogendyk, L., Wright, S.C., Lubensky, M., & Winnifred, L.R. (2016). Acting in solidarity: Cross-group contact between disadvantaged group members and advantaged group allies. *Journal of Social Issues, 72,* 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*, 637-652. https://doi.org/10.1111/josi.12083

Drury, J., & Reicher, S. (1999). The intergroup dynamics of collective empowerment: Substantiating the Social Identity Model of crowd behaviour. *Group Processes and Intergroup Relations, 2,* 381-402. https://doi.org/10.1177%2F1368430299024005.

Drury, J., & Reicher, S. (2000). Collective action and psychological change: The emergence of new social identities. *British Journal of Social Psychology*, *39*, 579-604. https://doi.org/10.1348/014466600164642.

Drury, J., & Reicher, S. (2009). Collective psychological empowerment as a model of social change: Researching crowds and power. *Journal of Social Issues*, *65*, 707-725. https://doi.org/10.1111/j.1540-4560.2009.01622.x.

Durrheim, K., Quayle, M., & Dixon, J. (2016). The struggle for the nature of "prejudice": "Prejudice" expression as identity performance. *Political Psychology, 37,* 17-35. https://doi.org/10.1111/pops.12310.

Elder, L., Greene, S., & Lizotte, M. K. (2021). Feminist and anti-feminist identification in the 21st century United States. *Journal of Women, Politics & Policy*, *42*, 243–259. https://doi.org/10.1080/1554477X.2021.1929607

Ellemers, N., & Barreto, M. (2003). The impact of relative group status: Affective, perceptual, and behavioural consequences. *Blackwell handbook of social psychology: Intergroup processes*, 325–343. https://doi.org/10.1002/9780470693421.ch16

Faludi, S. (1991). *Backlash: The undeclared war against American women*. New York: Crown

Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology*, *73*, 31–044. https://doi.org/10.1037/0022-3514.73.1.31

Fielding, K. S., & Hornsey, M. J. (2016). A social identity analysis of climate change and environmental attitudes and behaviors: Insights and opportunities. *Frontiers in Psychology*, *7*, 121. https://doi.org 10.3389/fpsyg.2016.00121/full

Fileborn, B., & Phillips, N. (2019). From 'Me Too' to 'Too far'? Contesting the boundaries of sexual violence in contemporary activism. In *#MeToo and the politics of social change* (pp. 99–115). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-15213-0_7

Fisher, M. L., & Exline, J. J. (2006). Self-forgiveness versus excusing: The roles of remorse, effort, and acceptance of responsibility. *Self and Identity*, *5*, 127–146. https://doi.org/10.1080/15298860600586123

Flood, M. (2019). Men and #MeToo: Mapping men's responses to anti-violence advocacy. In *#MeToo and the politics of social change* (pp. 285-300). Palgrave Macmillan, Cham.

Flood, M., Dragiewicz, M., & Pease, B. (2021). Resistance and backlash to gender

    equality. *Australian Journal of Social Issues*, *56*, 393–408.

    https://doi.org/10.1002/ajs4.137

Foster, M. D., & Rathlin, J. (2022). #MeToo as an 'angry mob' or in search of meaning?

    Using language to assess the focus of #MeToo tweets across four events. *Computers*

    in Human Behavior Reports, *5*, 1–9. https://doi.org/10.1016/j.chbr.2022.100173

Gee, A., Khalaf, A., & McGarty, C. (2007). Using group-based interaction to change

    stereotypes about people with mental disorders. *Australian Psychologist*, *42*(2), 98-

    105.

Gleeson, J., & Turner, B. (2019). Online feminist activism as performative consciousness-

    raising: A #MeToo case study. In *#MeToo and the politics of social change* (pp. 53–

    69). Palgrave Macmillan, Cham.

Gomez, E. M., & Kaiser, C. R. (2019). From pixels to protest: Using the Internet to confront

    bias at the societal level. In *Confronting prejudice and discrimination* (pp. 319–335).

    Academic Press. https://doi.org/10.1016/B978-0-12-814715-3.00011-4

Gómez, J.M., & Gobin, R.L. (2020). Black Women and Girls & #MeToo: Rape, Cultural

    Betrayal, & Healing. *Sex Roles, 82,* 1–12. https://doi.org/10.1007/s11199-019-01040-

    0

Goodman, D. (2001). *Promoting diversity and social justice: educating people from*

    privileged groups. SAGE Publications, Inc. https://doi.org/10.4135/9781452220468

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets

    of moral foundations. *Journal of Personality and Social Psychology, 96*, 1029–

    1046. https://doi.org/10.1037/a0015141

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology* (Vol. 47, pp. 55–130). Academic Press.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*, 366–385. https://doi.org/10.1037/a0021847

Greijdanus, H., de Matos Fernandes, C. A., Turner-Zwinkels, F., Honari, A., Roos, C. A., Rosenbusch, H., & Postmes, T. (2020). The psychology of online activism and social movements: Relations between online and offline collective action. *Current Opinion in Psychology*, *35*, 49–54. https://doi.org/10.1016/j.copsyc.2020.03.003

Griffiths, T., & Steyvers, M. (2004). Finding Scientific Topics. *Proceedings of the National Academy of Sciences, 101*, 5228-5235.

Gruber, A. (2009). Rape, feminism, and the war on crime. *Washington Law Review, 84,* 581–660.

Gunn, G. R., & Wilson, A. E. (2011). Acknowledging the skeletons in our closet: The effect of group affirmation on collective guilt, collective shame, and reparatory attitudes. *Personality and Social Psychology Bulletin*, *37*, 1474–1487. https://doi.org/10.1177/0146167211413607

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*, 98–116. https://doi.org/10.1007/s11211-007-0034-z

Hall, J. H., & Fincham, F. D. (2005). Self–forgiveness: The stepchild of forgiveness research. *Journal of Social and Clinical Psychology*, *24*, 621–637. https://doi.org/10.1521/jscp.2005.24.5.621

Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parcelling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, *2*, 233–256. https://doi.org/10.1177/109442819923002

Haslam, S. A., Oakes, P. J., Reynolds, K. J., & Turner, J. C. (1999). Social identity salience and the emergence of stereotype consensus. *Personality and Social Psychology Bulletin, 25,* 809–818. https://doi.org/10.1177/0146167299025007004

Haslam, S. A., Reicher, S. D., & Platow, M. J. (2015). Leadership: Theory and practice. In *APA Handbook of personality and social psychology, Volume 2: Group processes.* (pp. 67–94). American Psychological Association. https://doi.org/10.1037/14342-003

Haslam, S. A., Reicher, S. D., & Platow, M. J. (2020). *The new psychology of leadership: Identity, influence and power*. Routledge.

Haslam, S. A., Steffens, N. K., Peters, K., Boyce, R. A., Mallett, C. J., & Fransen, K. (2017). A social identity approach to leadership development: The 5R program. *Journal of Personnel Psychology*, *16*, 113-124. https://doi.org/10.1027/1866-5888/a000176.

Haslam, S. A., Turner, J. C., Oakes, P. J., Reynolds, K. J., Eggins, R. A., Nolan, M., & Tweedie, J. (1998). When do stereotypes become really consensual? Investigating the group-based dynamics of the consensualization process. *European Journal of Social Psychology*, *28*, 755–776.

Hässler, T., Ullrich, J., Sebben, S., Shnabel, N., Bernardino, M., Valdenegro, D., Van Laar, C., González, R., Visintin, E. P., Tropp, L. R., Ditlmann, R. K., Abrams, D., Aydin, A. L., Pereira, A., Selvanathan, H. P., von Zimmermann, J., Lantos, N. A., Sainz, M., Glenz, A., . . . Pistella, J. (2022). Need satisfaction in intergroup contact: A multinational study of pathways toward social change. *Journal of Personality and Social Psychology, 122,* 634–658. https://doi.org/10.1037/pspi0000365

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis, Second Edition: A Regression-Based Approach*. Guilford Publications.

Helm, B., Scrivens, R., Holt, T. J., Chermak, S. M., & Frank, R. (2022). Examining incel subculture on reddit. *Journal of Crime and Justice*, 1–19. https://doi.org/10.1080/0735648X.2022.2074867

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*, 29-29. https://doi.org/10.1038/466029a.

Hitch, G. (2021). *Defence Minister Linda Reynolds retracts 'lying cow' comment towards Brittany Higgins*. https://www.abc.net.au/news/2021-03-12/linda-reynolds-retracts-lying-cow-comment-brittany-higgins/13242902.

Hodson, G., Earle, M., & Craig, M. A. (2022). Privilege lost: How dominant groups react to shifts in cultural primacy and power. *Group Processes & Intergroup Relations*, *25*, 625–641. https://doi.org/10.1177/13684302211070524

Holt, T. J., Freilich, J. D., & Chermak, S. M. (2022). Examining the online expression of ideology among far-right extremist forum users. *Terrorism and Political Violence, 34,* 364-384. https://doi.org/10.1080/09546553.2019.1701446.

Hooks, B. (1995). Killing rage: Ending racism. New York, NY: Henry Holt & Company.

Howard, J. A., & Renfrow, D. G. (2014). Intersectionality. In *Handbook of the social psychology of inequality* (pp. 95-121). Springer, Dordrecht. https://doi.org/10.1111/bjso.12458

Hu, L., Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal,* 6, 1–55. https://doi.org/10.1080/10705519909540118

Iyer, A. (2022). Understanding advantaged groups' opposition to diversity, equity, and

inclusion (DEI) policies: The role of perceived threat. *Social and Personality Psychology Compass*, *16*, 1–14. https://doi.org/10.1111/spc3.12666

Iyer, A., & Leach, C.W. (2010). Helping disadvantaged out-groups challenge unjust

inequality: The role of group-based emotions. In S. Stürmer & M. Snyder (Eds.), *The psychology of prosocial behavior: Group processes, intergroup relations, and helping* (pp. 337–353). Wiley-Blackwell. https://doi.org/10.1002/9781444307948.ch17.

Iyer, A., & Ryan, M. K. (2009). Why do men and women challenge gender discrimination in

the workplace? The role of group status and in-group identification in predicting pathways to collective action. *Journal of Social Issues*, *65*, 791–814. https://doi.org/10.1111/j.1540-4560.2009.01625.x

Jaffe, A. E., Cero, I., & DiLillo, D. (2021). The #MeToo movement and perceptions of sexual

assault: College students' recognition of sexual assault experiences over time. *Psychology of Violence, 11*, 209–218. https://doi.org/10.1037/vio0000363

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the

production of false consciousness. *British Journal of Social Psychology*, *33*, 1–27. https://doi.org/10.1111/j.2044-8309.1994.tb01008.x

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory:

Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*, 881–919. https://doi.org/10.1111/j.1467-9221.2004.00402.x

Jost, J. T., Becker, J., Osborne, D., & Badaan, V. (2017). Missing in (collective) action:

Ideology, system justification, and the motivational antecedents of two types of

protest behaviour. *Current Directions in Psychological Science, 26*, 99–108. https://doi.org/10.1177/0963721417690633.

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

Kende, A., Nyúl, B., Lantos, N. A., Hadarics, M., Petlitski, D., Kehl, J., & Shnabel, N. (2020). A needs-based support for #MeToo: Power and morality needs shape women's and men's support of the campaign. *Frontiers in Psychology*, *11*, 1–19. https://doi.org/10.3389/fpsyg.2020.00593.

Kende, J., Phalet, K., Van den Noortgate, W., Kara, A., & Fischer, R. (2018). Equality revisited: A cultural meta-analysis of intergroup contact and prejudice. *Social Psychological and Personality Science*, *9*, 887–895. https://doi.org/10.1177/1948550617728993

Killian, L. M. (1985). The stigma of race: Who now bears the mark of Cain? *Symbolic Interaction*, *8*, 1–14. https://doi.org/10.1525/si.1985.8.1.1

Klein, O., Spears, R., & Reicher, S. (2007). Social Identity Performance: Extending the Strategic Side of SIDE. *Personality and Social Psychology Review, 11*, 28-45. https://doi.org/10.1177/1088868306294588.

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*, 594–609. doi.org/10.1177/1745691614554658

Kunst, J. R., Bailey, A., Prendergast, C., & Gundersen, A. (2019). Sexism, rape myths and feminist identification explain gender differences in attitudes toward the #Metoo

social media campaign in two countries. *Media Psychology*, *22*, 818–843. https://doi.org/10.1080/15213269.2018.1532300

Lanius, C. (2019). Torment porn or feminist witch hunt: Apprehensions about the #MeToo movement on /r/AskReddit. *Journal of Communication Inquiry*, *43*, 415–436. https://doi.org/10.1177/0196859919865250

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

Leach, C. W., Iyer, A., & Pedersen, A. (2007b). Angry opposition to government redress: When the structurally advantaged perceive themselves as relatively deprived. *British Journal of Social Psychology, 46*, 191–204. https://doi.org/10.1348/014466606X99360

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

Lee, B. (2021). The revolution will be organised. *The Saturday Paper*. https://www.thesaturdaypaper.com.au/opinion/topic/2021/12/18/the-revolution-will-be-organised/163974600013100#mtr

Leidner, B., Castano, E., Zaiser, E., & Giner-Sorolla, R. (2010). Ingroup glorification, moral disengagement, and justice in the context of collective violence. *Personality and Social Psychology Bulletin*, *36*, 1115–1129. https://doi.org/10.1177/0146167210376391

Leung, R., & Williams, R. (2019). #MeToo and intersectionality: An examination of the #MeToo movement through the R. Kelly scandal. *Journal of Communication Inquiry*, *43*, 349–371. https://doi.org/10.1177/0196859919874138

Lisak, D., Gardinier, L., Nicksa, S. C., & Cote, A. M. (2010). False allegations of sexual assault: An analysis of ten years of reported cases. *Violence Against Women*, *16*, 1318–1334. https://doi.org/10.1177/1077801210387747

Lisnek, J. A., Wilkins, C. L., Wilson, M., & Ekstrom, P. D. (2022). Backlash against the #MeToo movement: How women's voice causes men to feel victimized. *Group Processes & Intergroup Relations*, *25*, 682–702. https://doi.org/10.1177%2F13684302211035437

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioural sciences. *Behavior Research Methods, 49,* 433–442. https://doi.org/10.3758/s13428-016-0727-z

Loney-Howes, R. (2018). Shifting the rape script: "Coming out" online as a rape victim. *Frontiers: A Journal of Women Studies, 39*, 26–57. http://doi.org/10.5250/fronjwomestud.39.2.0026

Louis, W. R. (2009). Collective action—and then what? *Journal of Social Issues*, *65*, 727–748. https://doi.org/10.1111/j.1540-4560.2009.01623.x

Louis, W. R., Thomas, E., Chapman, C. M., Achia, T., Wibisono, S., Mirnajafi, Z., & Droogendyk, L. (2019). Emerging research on intergroup prosociality: Group members' charitable giving, positive contact, allyship, and solidarity with others. *Social and Personality Psychology Compass*, *13*, 1–16. https://doi.org/10.1111/spc3.12436

Lüders, A., Dinkelberg, A., & Quayle, M. (2022). Becoming "us" in digital spaces: How online users creatively and strategically exploit social media affordances to build up

socialidentity. *Acta Psychologica*, *228*, 103643.

https://doi.org/10.1016/j.actpsy.2022.103643

MacKinnon, D. P. (2012). *Introduction to statistical mediation analysis*. Routledge.

Mansbridge, J., & Shames, S. L. (2008). Toward a theory of backlash: Dynamic resistance and the central role of power. *Politics & Gender*, *4*, 623–634.

https://doi.org/10.1017/S1743923X08000500

Marques, J., Abrams, D., & Serôdio, R. G. (2001). Being better by being right: Subjective group dynamics and derogation of in-group deviants when generic norms are undermined. *Journal of Personality and Social Psychology, 81*, 436–447. https://doi.org/10.1037/0022-3514.81.3.436

McGarty, C., Bliuc, A. M., Thomas, E. F., & Bongiorno, R. (2009). Collective action as the material expression of opinion-based group membership. *Journal of Social Issues*, *65*, 839–857. https://doi.org/10.1111/j.1540-4560.2009.01627.x

McGarty, C., Thomas, E.F., Lala, G., Smith, L., & Bliuc, A. (2014). New technologies, new identities, and the growth of mass opposition in the Arab Spring. *Political Psychology, 35*, 725-740. https://doi.org/10.1111/pops.12060.

McQueen, A., & Klein, W. M. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity*, *5*, 289–354.

https://doi.org/10.1080/15298860600805325

Mendes, K., & Ringrose, J. (2019). Digital feminist activism: #MeToo and the everyday experiences of challenging rape culture. In *#MeToo and the politics of social change* (pp. 37–51). Palgrave Macmillan, Cham.

Mikołajczak, G., Becker, J. C., & Iyer, A. (2022). Women who challenge or defend the status quo: Ingroup identities as predictors of progressive and reactionary collective

action. *European Journal of Social Psychology*, 1–16.

https://doi.org/10.1002/ejsp.2842

Morgenroth, T., & Ryan, M. K. (2018). Addressing gender inequality: Stumbling blocks and

roads ahead. *Group Processes & Intergroup Relations*, *21*, 671-677.

https://doi.org/10.1177/1368430218786079

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample

size and determine power. *Structural Equation Modeling*, *9*, 599–620.

https://doi.org/10.1207/S15328007SEM0904_8

Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2017). *Regression and mediation analysis

using Mplus*. Los Angeles, CA: Muthén & Muthén.

Myers, D. G., & Bishop, G. D. (1970). Discussion effects on racial attitudes. *Science*, *169*,

778–779. doi:10.1126/science.169.3947.778

Nadler, A., & Shnabel, N. (2015). Intergroup reconciliation: Instrumental and socio-

emotional processes and the needs-based model. *European Review of Social

Psychology*, *26*, 93–125. https://doi.org/10.1080/10463283.2015.1106712

Nair, R., & Vollhardt, J. R. (2020). Intersectionality and relations between oppressed groups:

Intergroup implications of beliefs about intersectional differences and

commonalities. *Journal of Social Issues*, *76*, 993-1013.

https://doi.org/10.1111/josi.12409.

Nau, C., Zhang, J., Quan-Haase, A., & Mendes, K. (2022). Vernacular practices in digital

feminist activism on twitter: deconstructing affect and emotion in the #MeToo

movement. *Feminist Media Studies*, 1–17.

https://doi.org/10.1080/14680777.2022.2027496

Nicholas, L., & Agius, C. (2017). *The persistence of global masculinism: Discourse, gender

and neo-colonial re-articulations of violence*. Springer.

Nikita, M. (2016). Ldatuning: Tuning of the latent dirichlet allocation models parameters. Retrieved from https://CRAN.R-project.org/package=ldatuning.

Noor, M., Brown, R., & Prentice, G. (2008). Precursors and mediators of intergroup reconciliation in Northern Ireland: A new model. *British Journal of Social Psychology*, *47*, 481–495. https://doi.org/10.1348/014466607X238751

Noor, M., Shnabel, N., Halabi, S., & Nadler, A. (2012). When suffering begets suffering: The psychology of competitive victimhood between adversarial groups in violent conflicts. *Personality and Social Psychology Review*, *16*, 351–374. https://doi.org/10.1177/1088868312440048

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*, 215–218. https://doi.org/10.1177/1745691611406922

Okuyan, M., & Vollhardt, J. (2022). The role of group versus hierarchy motivations in dominant groups' perceived discrimination. *Group Processes & Intergroup Relations, 25,* 54–80. https://doi.org/10.1177/13684302211053543

Osborne, D., Jost, J. T., Becker, J. C., Badaan, V., & Sibley, C. G. (2019). Protesting to challenge or defend the system? A system justification perspective on collective action. *European Journal of Social Psychology*, *49*, 244–269. https://doi.org/10.1002/ejsp.2522

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, *66*, 1025–1060. https://doi.org/10.1111/1467-6494.00041

Pedahzur, A., & Canetti-Nisim, D. (2004). Support for right-wing extremist ideology: socio-economic indicators and socio-psychological mechanisms of social identification. *Comparative Sociology, 3,* 1-36. https://doi.org/10.1163/1569133041513756.

Peetz, J., Gunn, G. R., & Wilson, A. E. (2010). Crimes of the past: Defensive temporal distancing in the face of past in-group wrongdoing. *Personality and Social Psychology Bulletin*, *36*, 598–611. https://doi.org/10.1177/0146167210364850

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Perez, M. J., & Salter, P. S. (2020). A critical race reading of collective victimhood. In J. R. Vollhardt (Ed.), *The social psychology of collective victimhood* (pp. 319–336). Oxford University Press.

PettyJohn, M. E., Muzzey, F. K., Maas, M. K., & McCauley, H. L. (2019). #HowIWillChange: Engaging men and boys in the #MeToo movement. *Psychology of Men & Masculinities*, *20*, 612–622. https://doi.org/10.1037/men0000186

Platow, M. J., Van Rooy, D., Augoustinos, M., Spears, R., Bar Tal, D., & Grace, D. M. (2019). Prejudice is about collective values, not a biased psychological system. *New Zealand Journal of Psychology, 47,* 15-22.

Postmes, T., & Spears, R. (1998). Deindividuation and anti-normative behavior: A meta-analysis. *Psychological Bulletin*, *123*, 238–259. doi:10.1037/0033- 2909.123.3.238

Postmes, T., Haslam, S. A., & Swaab, R. I. (2005). Social influence in small groups: An interactive model of social identity formation. *European Review of Social Psychology*, *16*, 1–42. https://doi.org/10.1080/10463280440000062

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891. https://doi.org/10.3758/BRM.40.3.879

Prentice, D., & Paluck, E. L. (2020). Engineering social change using social norms: Lessons from the study of collective action. *Current Opinion in Psychology*, *35*, 138-142. https://doi.org/10.1016/j.copsyc.2020.06.012

Rafail, P., & Freitas, I. (2019). Grievance articulation and community reactions in the men's rights movement online. *Social Media & Society*, *5*, 1–11. https://doi.org/10.1177/2056305119841387

RAINN (2015). *Statistics.* https://www.rainn.org/statistics/victims-sexual-violence

Reicher, S. D. (1996). 'The Battle of Westminster': Developing the social identity model of crowd behaviour in order to explain the initiation and development of collective conflict. *European Journal of Social Psychology, 26*, 115-134.

Reicher, S. D. (2004). The context of social identity: Domination, resistance and change. *Political Psychology, 25*, 921-945. https://doi.org/10.1111/j.1467-9221.2004.00403.x.

Reicher, S., Haslam, S. A., & Hopkins, N. (2005). Social identity and the dynamics of leadership: Leaders and followers as collaborative agents in the transformation of social reality. *The Leadership Quarterly*, *16*, 547-568

Reicher, S., & Hopkins, N. (1996). Seeking influence through characterizing self-categories: An analysis of anti-abortionist rhetoric. *British Journal of Social Psychology, 35,* 297–311. https://doi.org/10.1111/j.2044-8309.1996.tb01099.x.

Reicher, S. D., Haslam, S. A., & Platow, M. J. (2007). The new psychology of leadership. *Scientific American Mind*, *18*, 22-29.

Reicher, S. D., Spears, R., & Haslam, A. S. (2010). The social identity approach in social psychology. In M. Wetherell, & C. T. Mohanty (Eds.) *Sage Identities Handbook* (pp. 45-62). London: Sage. http://dx.doi.org/10.4135/9781446200889.n4.

Reicher, S. D., Spears, R., & Postmes, T. (1995). A social identity model of deindividuation phenomena. *European Review of Social Psychology*, *6*, 161–198. https://doi.org/10.1080/14792779443000049

Reicher, S., & Ulusahin, Y. (2020). Resentment and Redemption: On the Mobilization of Dominant Group Victimhood. In J. R. Vollhardt (Ed.), *The social psychology of collective victimhood* (pp. 275-294). Oxford University Press.

Richard, F. D., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*, 331–363. https://doi.org/10.1037/1089-2680.7.4.331

Ringrose, J., & Lawrence, E. (2018). Remixing misandry, manspreading, and dick pics: Networked feminist humour on Tumblr. *Feminist Media Studies, 18,* 686–704. https://doi.org/10.1080/14680777.2018.1450351

Rivera-Rodriguez, A., Larsen, G., & Dasgupta, N. (2022). Changing public opinion about gender activates group threat and opposition to feminist social movements among men. *Group Processes & Intergroup Relations*, *25*, 811–829. https://doi.org/10.1177/13684302211048885

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58,* 1064-1082.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, *91*, 1–40. https://doi.org/10.18637/jss.v091.i02

Rotella, K. N., & Richeson, J. A. (2013). Motivated to "forget" the effects of in-group wrongdoing on memory and collective guilt. *Social Psychological and Personality Science*, *4*, 730–737. https://doi.org/10.1177/1948550613482986

Roth-Cohen, O. (2021). Viral feminism: #MeToo networked expressions in feminist Facebook groups. *Feminist Media Studies*, 1–17. https://doi.org/10.1080/14680777.2021.1906295

Russell, G. M. (2011). Motives of heterosexual allies in collective action for equality. *Journal of Social Issues, 67,* 376–393. https://doi.org/10.1111/j.1540-4560.2011.01703.x.

Ryan, M. (2022). To advance equality for women, use the evidence. *Nature, 604,* 403-403. https://doi.org/10.1038/d41586-022-01045-y

Saab, R., Tausch, N., Spears, R., & Cheung, W. Y. (2015). Acting in solidarity: Testing an extended dual pathway model of collective action by bystander group members. *British Journal of Social Psychology*, *54*, 539–560. https://doi.org/10.1111/bjso.12095

Schmidt, A. L., Zollo, F., Scala, A., Betsch, C., & Quattrociocchi, W. (2018). Polarization of the vaccination debate on Facebook. *Vaccine*, *36*, 3606–3612. https://doi.org/10.1016/j.vaccine.2018.05.040

Schmitz, R. M., & Kazyak, E. (2016). Masculinities in Cyberspace: An Analysis of Portrayals of Manhood in Men's Rights Activist Websites. *Social Sciences*, *5*, 1-16. https://doi.org/10.3390/socsci5020018

Schneider, K. T., & Carpenter, N. J. (2019). Sharing #MeToo on Twitter: Incidents, coping responses, and social reactions. *Equality, Diversity and Inclusion: An International Journal, 39*, 87-100. https://doi.org/10.1108/EDI-09-2018-0161.

Segerberg, A., & Bennett, W. L. (2011). Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests. *The Communication Review*, *14*, 197–215. https://doi.org/10.1080/10714421.2011.597250

Selvanathan, H. P., & Jetten, J. (2020). From marches to movements: Building and sustaining a social movement following collective action. *Current Opinion in Psychology, 35,* 81–85. https://doi.org/10.1016/j.copsyc.2020.04.004.

Selvanathan, H. P., Lickel, B., & Dasgupta, N. (2020a). An integrative framework on the impact of allies: How identity-based needs influence intergroup solidarity and social

movements. *European Journal of Social Psychology*, *50*, 1344-1361.

https://doi.org/10.1002/ejsp.2697

Selvanathan, H. P., Lickel, B., & Jetten, J. (2020b). Collective psychological ownership and

the rise of reactionary counter-movements defending the status quo. *British Journal of

Social Psychology*, *60*, 587–609. https://doi/full/10.1111/bjso.12418

Settles, I. H., & Buchanan, N. T. (2014). Multiple groups, multiple identities, and

intersectionality. In V. Martínez & Y.-Y. Hong (Eds.), The Oxford *Handbook of

multicultural identity* (pp. 160–180). Oxford University Press.

https://doi.org/10.1093/oxfordhb/9780199796694.013.017.

Sheeran, P., & Webb, T. L. (2016). The intention–behavior gap. *Social and Personality

Psychology Compass, 10, 503*-518. https://doi.org/10.1111/spc3.12265

Shepherd, L., Fasoli, F., Pereira, A., & Branscombe, N. R. (2018). The role of threat,

emotions, and prejudice in promoting collective action against immigrant groups.

European Journal of Social Psychology, *48*, 447-459.

https://doi.org/10.1002/ejsp.2346.

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation

theory. *Advances in Experimental Social Psychology*, *38*, 183–242.

https://doi.org/10.1016/S0065-2601(06)38004-5

Shnabel, N., & Nadler, A. (2008). A needs-based model of reconciliation: satisfying the

differential emotional needs of victim and perpetrator as a key to promoting

reconciliation. *Journal of Personality and Social Psychology, 94*, 116–132.

https://doi.org/10.1037/0022-3514.94.1.116

Shnabel, N., Halabi, S., & Noor, M. (2013a). Overcoming competitive victimhood and

facilitating forgiveness through re-categorization into a common victim or perpetrator

identity. *Journal of Experimental Social Psychology*, *49*, 867–877.

https://doi.org/10.1016/j.jesp.2013.04.007

Shnabel, N., Nadler, A., Ullrich, J., Dovidio, J. F., & Carmi, D. (2009). Promoting

reconciliation through the satisfaction of the emotional needs of victimized and

perpetrating group members: The needs-based model of reconciliation. *Personality*

and Social Psychology Bulletin, *35*, 1021–1030.

https://doi.org/10.1177/0146167209336610

Shnabel, N., Ullrich, J., Nadler, A., Dovidio, J. F., & Aydin, A. L. (2013b). Warm or

competent? Improving intergroup relations by addressing threatened identities of

advantaged and disadvantaged groups. *European Journal of Social Psychology*, *43*,

482–492. https://doi.org/10.1002/ejsp.1975

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy*

and oppression. New York, NY: Cambridge University Press.

Siem, B., Von 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*, 137–148. https://doi.org/10.1002/ejsp.1929

Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles

in R. *Journal of Open Source Software*, *37,* 1–3. https://doi.org/10.21105/joss.00037

SimanTov-Nachlieli, I., Shnabel, N., Aydin, A. L., & Ullrich, J. (2018). Agents of

prosociality: Agency affirmation promotes mutual prosocial tendencies and behavior

among conflicting groups. *Political Psychology*, *39*, 445–463.

https://doi.org/10.1111/pops.12418

Simon, B., & Klandermans, B. (2001). Politicized collective identity: A social psychological analysis. *American Psychologist, 56*, 319–331. https://doi.org/10.1037/0003-066X.56.4.319.

Sloan, L., Morgan, J., Housley, W., Williams, M., Edwards, A., Burnap, P., & Rana, O. (2013). Knowing the Tweeters: Deriving sociologically relevant demographics from Twitter. *Sociological Research Online*, *18*, 74–84. https://doi.org/10.5153%2Fsro.3001

Smith, L. G., Blackwood, L., & Thomas, E. F. (2020a). The need to refocus on the group as the site of radicalization. *Perspectives on Psychological Science*, *15*, 327-352. https://doi.org/10.1177/1745691619885870

Smith, L. G. E., & Postmes, T. (2009). Intra-group interaction and the development of norms which promote inter-group hostility. *European Journal of Social Psychology*, *39*, 130–144. https://doi.org/10.1002/ejsp.464

Smith, L. G. E., & Postmes, T. (2011). Shaping stereotypical behaviour through the discussion of social stereotypes. *British Journal of Social Psychology*, *50*, 74–98. https://doi.org/10.1348/014466610X500340

Smith, L. G. E., McGarty, C., & Thomas, E. F. (2018). After Aylan Kurdi: How tweeting about death, threat, and harm predict increased expressions of solidarity with refugees over time. *Psychological science*, *29*, 623-634. https://doi.org/10.1177/0956797617741107

Smith, L. G. E., Thomas, E. F., & McGarty, C. (2015). "We must be the change we want to see in the world": Integrating norms and identities through social interaction. *Political Psychology*, *36*, 543–557. https://doi.org/10.1111/pops.12180

Smith, L.G.E., Wakeford, L., Cribbin, T.F., Barnett, J., & Hou, W. K. (2020b). Detecting psychological change through mobilizing interactions and changes in extremist

linguistic style. *Computers in Human Behaviour, 108*, 1-49.

https://doi.org/10.1016/j.chb.2020.106298.

Stanley, S. K., Milfont, T. L., Wilson, M. S., & Sibley, C. G. (2019). The influence of social

dominance orientation and right-wing authoritarianism on environmentalism: A five-

year cross-lagged analysis. *Plos One*, *14*,

https://doi.org/10.1371/journal.pone.0219067

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the

self. *Advances in Experimental Social Psychology*, *21*, 261–302.

https://doi.org/10.1016/S0065-2601(08)60229-4

Stewart, A. L. (2017). Men's collective action willingness: Testing different theoretical

models of protesting gender inequality for women and men. *Psychology of Men &*

Masculinity, *18*, 372–381. https://doi.org/10.1037/men0000068

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In Landauer, T. K.,

McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.) *Handbook of latent semantic*

analysis (pp. 424-440). Psychology Press.

Stott, C., & Drury, J. (1999). The inter-group dynamics of collective empowerment: A social

identity model of crowd behavior. *Group Processes & Intergroup Relations*, *2*, 381-

402. https:// doi.org/10.1177/1368430299024005.

Strelan, P., Weick, M., & Vasiljevic, M. (2014). Power and revenge. *British Journal of Social*

Psychology, *53*, 521-540. https://doi.org/10.1111/bjso.12044

Subašić, E., Hardacre, S., Elton, B., Branscombe, N. R., Ryan, M. K., & Reynolds, K. J.

(2018). "We for She": Mobilising men and women to act in solidarity for gender

equality. *Group Processes & Intergroup Relations*, *21*, 707–724.

https://doi.org/10.1177/1368430218763272

Subašić, E., Reynolds, K. J., & Mohamed, M. S. (2014). Changing identities to change society: Leadership as a contest for influence and collective mobilization. In *Psychology of Change* (pp. 260-277). Psychology Press.

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*, *12*, 330–352. https://doi.org/10.1177/1088868308323223

Subašić, E., Reynolds, K.J, Turner, J.C., Veenstra, K.E., & Haslam, A. (2011). Leadership, power and the use of surveillance: Implications of shared social identity for leaders' capacity to influence. *The Leadership Quarterly, 22*, 170-181. https://doi.org/10.1016/j.leaqua.2010.12.014.

Sudkämper, A., Ryan, M. K., Kirby, T. A., & Morgenroth, T. (2020). A comprehensive measure of attitudes and behaviour: Development of the support for gender equality among men scale. *European Journal of Social Psychology*, *50*, 256–277. https://doi.org/10.1002/ejsp.2629

Sullivan, D., Landau, M. J., Branscombe, N. R., & Rothschild, Z. K. (2012). Competitive victimhood as a response to accusations of ingroup harm doing. *Journal of Personality and Social Psychology, 102*, 778–795. https://doi.org/10.1037/a0026573

Szekeres, H., Shuman, E., & Saguy, T. (2020). Views of sexual assault following #MeToo: The role of gender and individual differences. *Personality and Individual Differences*, *166*, 1–6. https://doi.org/10.1016/j.paid.2020.110203

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, *33*, 1-39.

Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. In W. Austin & S. Worchel (Eds), *The social psychology of intergroup relations*. Monterey, CA: Brookes/Cole.

Tausch, N., Becker, J. C., Spears, R., Christ, O., Saab, R., Singh, P., & Siddiqui, R. N. (2011). Explaining radical group behavior: Developing emotion and efficacy routes to normative and nonnormative collective action. *Journal of Personality and Social Psychology, 101*, 129–148. https://doi.org/10.1037/a0022728

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

Terry, D. J., & Hogg, M. A. (1996). Group norms and the attitude-behavior relationship: A role for group identification. *Personality and social psychology bulletin*, *22*, 776–793. https://doi.org/10.1177/0146167296228002

Thomas, E. F., & McGarty, C. (2009). The role of efficacy and moral outrage norms in creating the potential for international development activism through group-based interaction. *British Journal of Social Psychology*, *48*, 115–134. https://doi.org/10.1348/01446 6608X313774

Thomas, E. F., Bury, S. M., Louis, W. R., Amiot, C. E., Molenberghs, P., Crane, M. F., & Decety, J. (2019). Vegetarian, vegan, activist, radical: Using latent profile analysis to examine different forms of support for animal welfare. *Group Processes & Intergroup Relations*, *22*, 836-857. https://doi.org/10.1177/1368430218824407.

Thomas, E. F., Duncan, L., McGarty, C., Louis, W. R., & Smith, L. G. (2022a). MOBILISE: A Higher-Order Integration of Collective Action Research to Address Global Challenges. *Political Psychology*, 1–58. https://doi.org/10.1111/pops.12811

Thomas, E. F., Leggett, N., Kernot, D., Mitchell, L., Magsarjav, S., & Weber, N. (2022b). Reclaim the beach: How offline events shape online interactions and networks amongst those who support and oppose right-wing protest. *Studies in Conflict & Terrorism*, 1-23. https://doi.org/10.1080/1057610X.2022.2034222

Thomas, E.F., Mavor, K., & McGarty, C. (2012). Social identities facilitate and encapsulate action-relevant constructs: A test of the social identity model of collective action. *Group Processes & Intergroup Relations, 15,* 75-88. https://doi.org/10.1177%2F1368430211413619.

Thomas, E. F., McGarty, C., & Mavor, K. I. (2016). Group interaction as the crucible of identity formation: A glimpse at the origins of collective identity and action. *Group Processes & Intergroup Relations*, *19*, 137–151. https://doi.org/10.1177/1368430215612217

Thomas, E. F., Zubielevitch, E., Sibley, C. & Osborne, D. (2020). Testing the social identity model of collective action longitudinally and across structurally disadvantaged and advantaged groups. *Personality and Social Psychology Bulletin, 46,* 823–838. https://doi.org/10.1177/0146167219879111.

Trott, V. (2021). Networked feminism: counterpublics and the intersectional issues of #MeToo. *Feminist Media Studies, 21,* 1125–1142. https://doi.org/10.1080/14680777.2020.1718176

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

Turner-Zwinkels, F. M., & van Zomeren, M. (2021). Identity expression through collective action: How identification with a politicized group and its identity contents differently motivated identity-expressive collective action in the U.S. 2016 presidential elections.

Personality and Social Psychology Bulletin, 47, 499–513.

https://doi.org/10.1177/0146167220933406.

Ullman, S. E., & Townsend, S. M. (2007). Barriers to working with sexual assault survivors: A qualitative study of rape crisis center workers. *Violence Against Women, 13*, 412–443. https://doi.org/10.1177%2F1077801207299191

Uluğ, Ö. M., Chayinska, M., & Tropp, L. R. (2022). Conceptual, methodological, and contextual challenges in studying collective action: Recommendations for future research. *TPM: Testing, Psychometrics, Methodology in Applied Psychology*, *29*, 9–22. https://doi.org/10.4473/TPM29.1.2

Ussher, J. M., Hawkey, A., Perz, J., Liamputtong, P., Sekar, J., Marjadi, B., ... & Brook, E. (2022). Crossing boundaries and fetishization: Experiences of sexual violence for trans women of color. *Journal of Interpersonal Violence*, *37*, 3552–3584. https://doi.org/10.1177/0886260520949149

van Haperen, S., Uitermark, J., & Nicholls, W. (2022). The Swarm versus the Grassroots: places and networks of supporters and opponents of Black Lives Matter on Twitter. *Social Movement Studies*, 1–19. https://doi.org/10.1080/14742837.2022.2031954

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

van Zomeren, M., Spears, R., Fischer, A. H., & Leach, C. W. (2004). Put your money where your mouth is! Explaining collective action tendencies through group-based anger and group efficacy. *Journal of Personality and Social Psychology, 87*, 649–664. https://doi.org/10.1037/0022-3514.87.5.649.

Vergani, M., & Bliuc, A. M. (2015). The evolution of the ISIS' language: a quantitative analysis of the language of the first year of Dabiq magazine. *Sicurezza, Terrorismo e Societa*, *2*, 7–20.

Vestergen, S., Drury, J., & Hammar Chiriac, E. (2019). How participation in collective action changes relationships, behaviours, and beliefs: An interview study of the role of inter- and intragroup processes. *Journal of Social and Political Psychology, 7,* 76-99. https://doi.org/10.5964/jspp.v7i1.903.

Vollhardt, J. (2020). *The social psychology of collective victimhood*. Oxford University Press.

Vollhardt, J. R., Mazur, L. B., & Lemahieu, M. (2014). Acknowledgment after mass violence: Effects on psychological well-being and intergroup relations. *Group Processes & Intergroup Relations*, *17*, 306–323. https://doi.org/10.1177/1368430213517270

Wang, Y. A., & Rhemtulla, M. (2021). Power analysis for parameter estimation in structural equation modelling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science*, *4*, 1–17. https://doi.org/10.1177/2515245920918253.

Weiser, D. A. (2017). Confronting myths about sexual assault: A feminist analysis of the false report literature. *Family Relations*, *66*, 46–60. https://doi.org/10.1111/fare.12235

West, K., Greenland, K., & van Laar, C. (2021). Implicit racism, colour blindness, and narrow definitions of discrimination: Why some White people prefer 'All Lives Matter' to 'Black Lives Matter'. *British Journal of Social Psychology*, *60*, 1136-1153.

White, F., Maunder, R., Verrelli, S. (2020). Text-based E-contact: Harnessing cooperative Internet interactions to bridge the social and psychological divide. *European Review of Social Psychology*, *31,* 76-119. https://doi.org/10.1080/10463283.2020.1753459.

White, F., Newson, M., Verrelli, S., Whitehouse, H. (2021). Pathways to prejudice and outgroup hostility: Group alignment and intergroup conflict among football

fans. *Journal of Applied Social Psychology, 51,* 660-666.

https://doi.org/10.1111/jasp.12773.

White, F., Turner, R., Verrelli, S., Harvey, L., Hanna, J. (2019). Improving intergroup

relations between Catholics and Protestants in Northern Ireland via E-

contact. *European Journal of Social Psychology, 49,* 429-438.

https://doi.org/10.1002/ejsp.2515.

Wilkins, C. L., Wellman, J. D., & Schad, K. D. (2017). Reactions to anti-male sexism claims:

The moderating roles of status-legitimizing belief endorsement and group

identification. *Group Processes & Intergroup Relations*, *20*, 173–185.

https://doi.org/10.1177/1368430215595109

Wilkins, C. L., Wellman, J. D., Babbitt, L. G., Toosi, N. R., & Schad, K. D. (2015). You can

win but I can't lose: Bias against high-status groups increases their zero-sum beliefs

about discrimination. *Journal of Experimental Social Psychology*, *57*, 1–14.

https://doi.org/10.1016/j.jesp.2014.10.008

Wilkins, D. J., Livingstone, A. G., & Levine, M. (2019). Whose tweets? The rhetorical

functions of social media use in developing the Black Lives Matter movement. *British

Journal of Social Psychology*, *58*, 786–805. https://doi.org/10.1111/bjso.12318

Williams, J. B. (2021). Maximizing #MeToo: Intersectionality & the Movement. *Boston

College Law Review, 62*, 1797–1864.

Williams, T., Testa, P. F., Britzman, K., & Hibbing, M. V. (2021). Messengers Matter: Why

Advancing Gender Equity Requires Male Allies. *PS: Political Science & Politics, 54,*

512-513. https://doi.org/10.1017/S1049096521000093

Witt, T. (2020). 'If i cannot have it, i will do everything i can to destroy it.' The canonization

of Elliot Rodger: 'Incel' masculinities, secular sainthood, and justifications of

ideological violence. *Social Identities*, *26*, 675–689.

https://doi.org/10.1080/13504630.2020.1787132

Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of

competence or morality. *Journal of Personality and Social Psychology*, *67*, 222–232.

https://doi.org/10.1037/0022-3514.67.2.222

Woodyatt, L., & Wenzel, M. (2013). The psychological immune response in the face of

transgressions: Pseudo self-forgiveness and threat to belonging. *Journal of*

Experimental Social Psychology, *49*, 951–958.

https://doi.org/10.1016/j.jesp.2013.05.016

World Health Organization. (2021). Violence against women prevalence estimates, 2018:

global, regional, and national prevalence estimates for intimate partner violence

against women and global and regional prevalence estimates for non-partner sexual

violence against women.

Xing, Y., Wang, X., Qiu, C., Li, Y., & He, W. (2022). Research on opinion polarization by

big data analytics capabilities in online social networks. *Technology in Society, 68,* 1–

12. https://doi.org/10.1016/j.techsoc.2022.101902

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time

on Twitter. *Bulletin of Science, Technology & Society*, *30*, 316–327.

https://doi.org/10.1177%2F0270467610380011

Young, I. F., & Sullivan, D. (2016). Competitive victimhood: A review of the theoretical and

empirical literature. *Current Opinion in Psychology*, *11*, 30–34.

https://doi.org/10.1016/j.copsyc.2016.04.004

Yuan, K. H., & Chan, W. (2016). Measurement invariance via multigroup SEM: Issues and

solutions with chi-square-difference tests. *Psychological Methods*, *21*, 405–426.

https://psycnet.apa.org/doi/10.1037/met0000080

Zhou, S., Page-Gould, E., Aron, A., Moyer, A., & Hewstone, M. (2019). The extended

    contact hypothesis: A meta-analysis on 20 years of research. *Personality and Social*

    Psychology Review, 23, 132–160. https://doi.org/10.1177/1088868318762647

Appendix A

Study 3. Mitigated threat condition

NEWS

Men join campaign to end sexual harassment

In response to women sharing their experiences of sexual assault and harassment online, thousands of men have come forward to join the fight to end discrimination against women.

Many have posted about the actions they promise to take to stand up against sexism, from donating to women's shelters to calling out other men for engaging in sexist behaviour.

Christopher Kaczor
@chriskaczor
Follow

I promise to speak out when I see sexual harassment taking place

12:03 PM - 2 Sep 2018

82 Retweets 176 Likes

23   82   176

Brian Hedges
@b_hedges58
Follow

I will teach my sons to honor and respect women

4:52 PM - 2 Sep 2018

45 Retweets 102 Likes

Study 3. Mitigated threat control condition

NEWS

Men leading by example: In praise of Fathers

This Father's Day thousands of men have taken to social media to share their love of parenting and thank their own fathers for being positive male role models in their lives.

Many men posted about the joys of watching their children grow and succeed, while others thanked their own dads for their love and support.

Christopher Kaczor
@chriskaczor

Follow

My dad's provided for his family his whole life. He's the best male role model I know.

12:03 PM - 02 Sep 2018

82 Retweets 176 Likes

💬 23   🔁 82   ♡ 176   ✉

Brian Hedges
@b_hedges58

Follow

my father always let us know how much we were loved

4:52 PM - 02 Sep 2018

45 Retweets 102 Likes