Is Moral Disengagement Really Maximal Moral Engagement?

BY

ALLISON BETH MUELLER
B.A., Northwestern University, 2009
M.A., University of Illinois at Chicago, 2015

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Chicago, 2019

Chicago, Illinois

Defense Committee:

      Linda Skitka, Committee Chair & Advisor
      Allyson Holbrook
      Tomas Stahl
      Maja Kutlaca, University of Osnabrück
      Kimberly Quinn, DePaul University

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Many thanks to my advisor, Linda Skitka, for sharing her invaluable expertise with me throughout my graduate career. I would also like to thank my committee members for strengthening this work: Allyson Holbrook, Maja Kutlaca, Kimberly Quinn, and Tomas Stahl. Finally, I would like to express gratitude to my partner, family, and friends, all of whom supported me through every step of this journey and never stopped believing in me.

SUMMARY

The goal of this dissertation was to explore why people are willing to harm their political opponents. Although Bandura (1999) theorized that a variety of processes could license reprehensible behavior, I contended that *moral justification* (i.e., reconstruing harmful acts as supportive of a higher moral purpose) would be a primary path toward violence because it would preserve (or even enhance) people's positive self-views. To the extent that people construe morally motivated violence to be in service of the greater good at personal risk, they may: (1) downplay the transgressiveness of their behavior, (2) focus on the morally mandated outcome achieved, (3) ultimately construe the behavior to be morally upstanding, and (4) temporarily enjoy an inflated moral self-concept. Results did not support these hypotheses. People strategically endorsed mild harms at contentious political rallies to serve their strong moral convictions (i.e., picketing and staging sit-ins). But they resisted endorsing severe harms against obnoxious protesters (e.g., pushing, kicking, or spitting at those targets), especially when they had a strong moral commitment to the issue at stake. Moreover, morally motivated endorsement of mild harms was never linked to boosts in moral self-image. Overall, these findings suggest a limit to morally motivated political engagement: People feel comfortable endorsing mild forms of activism to serve their strong moral convictions but not violent forms. Future research employing a variety of methodologies (e.g., field and archival methods) should continue exploring the psychological levers that enable people to commit atrocities.

Is Moral Disengagement Really Maximal Moral Engagement?

Across the famous *Harry Potter* book series, wizards become increasingly embroiled in a war between good and evil. On the one side are the wizards who fight for inclusiveness and equality for all magical creatures, but on the other side are wizards fighting to re-establish tradition and a clear power hierarchy. Harry Potter is the poster child for the "good" side, and he is so extraordinary because he consistently risks everything to fight for what he believes is right. Over the course of the seven books, Harry risks his life several times to thwart evil teachers' sinister plans, breaks into the Ministry of Magic to uncover its corruption, and duels with the evilest wizard to ever live. To those on the "good" side, and to Harry himself, he is a hero.

To his enemies who wish to re-establish old-world tradition, however, Harry Potter is a menace. Harry has no regard for the law, even when his actions put others in danger; he seemingly has no issue with casting spells against anyone who disagrees with his vision for the future; he even used an unforgiveable curse to break into Gringotts Wizarding Bank. How could Harry, a hero by some standards, commit these harmful actions? It is possible that Harry may have felt empowered to aggress against his enemies because he was disengaged from normative standards. That is, Harry may have temporarily suspended his normal standards about not harming others to fight the "good" fight.

How could Harry Potter display—seemingly concurrently—extraordinary moral courage *and* deep moral disregard? The goal of my dissertation is to reconcile these two extremes of morality. Specifically, I contend that both moral courage and moral disregard could be driven by the same process: Strong moral conviction that a stance is right or wrong—i.e., moral or immoral—may make it easier to disengage from normative standards to serve that belief, including the decision to harm political opponents. For transgressors, disengagement from

normative standards (i.e., harming others) to serve a higher moral purpose might be experienced as a morally courageous action. The more one perceives that he or she has taken a stand for a cause at great perceived personal risk, the more likely he or she is to feel morally courageous and experience an inflated moral self-concept.

Going one step further, I propose that people are only willing to override the norm to not harm others when doing so does not threaten their moral self-concepts. Although Bandura (1999) theorized that there are also non-moral routes to moral disengagement, I argue that having compelling *moral* justification to harm others is the only route to inflicting harm on others that preserves (or even enhances) transgressors' moral integrity. Moralizing one's actions in these contexts is therefore a primary path towards morally disengaged actions. In contrast to Bandura (1999), I theorize that non-moral justifications will not sufficiently offset the negative impact of harmful acts on people's moral self-concepts, and they should therefore be unlikely to independently predict willingness to harm others for a non-moralized cause.

Before exploring these new predictions, I first review theory and research on morally motivated transgressions from transgressors' perspectives, with a particular emphasis on the role of strong moral convictions in these processes.

Morally Motivated Transgressions

People generally act in line with their moral standards. However, the inhibition of immoral impulses is an active process that requires self-regulation and effort; that is, self-regulatory mechanisms do not serve as internal regulators of moral behavior unless people choose to activate them. When people choose to stop self-regulating, they are said to be *disengaged from moral standards* (or, *morally disengaged* for short; Bandura, 1999).

When might people decide to disengage from their moral standards? Given that people experience self-condemnation when they disobey their own moral standards, they need compelling justification to negate any self-condemnation they typically feel. Bandura (1999) theorized that there are several justifications that enable people to engage in reprehensible conduct. They can (1) reconstrue immoral behavior in a positively biased way, so that it no longer is perceived as immoral; (2) highlight how their transgressions pale in comparison to the atrocities committed by others; and (3) use euphemistic language to label transgressions as something less serious than they actually are (see Bandura, 1999 for a review).

Of most relevance to my perspective on why people are willing to harm others is the tendency for people to reconstrue immoral behavior in a positively biased way. People may feel personally and socially entitled to transgress because they perceive that immoral behavior is required to achieve a higher moral purpose. To the extent that a transgression serves a noble cause, people may cognitively reconstrue that behavior to be not only positive, but as a moral imperative. When that cognitive reconstruction process occurs before transgressing, people no longer see themselves as immoral and worthy of condemnation for committing the action; they see themselves as agents of morality who are acting in the service of the greater good. Perhaps the best illustration of this process at work is in military conduct: To justify wartime violence, members of the military may see themselves as, "…fighting against ruthless oppressors, protecting their cherished values, preserving world peace, saving humanity from subjugation, or honoring their country's commitments…given people's dexterous facility for justifying violent means, all kinds of inhumanities get clothed in moral wrappings" (Bandura, 1999, p. 195). In short, people seem to be skilled at creatively reframing their immoral behaviors (e.g., killing,

bombing, attacking opponents) as supportive of a greater moral purpose (e.g., bolstering a moralized cause, protecting the free world, fighting the good fight).

To be sure, justifications like those described above shed light on why some unspeakable crimes against humanity take place, explaining how ordinary people can convince themselves that truly abhorrent actions are justified. Although I fully acknowledge and appreciate the knowledge gained from existing theories of moral disengagement, I posit that there remains a critical, yet unexplored, point of clarification that could help improve our understanding of moral disengagement.

Specifically, it is critical to revisit what it means to support or fail to support a "moral standard." People's use of the term "moral standard" could refer to *normative* standards—i.e., what society and culture deem as right or convention in a given context, but that in another context might be inappropriate. For example, it is "wrong" to wear a bathing suit to a dissertation defense, even if it is perfectly "right" to wear one to the beach. But people's use of the term "moral standard" could also refer to *personal* moral standards—i.e., beliefs about right or wrong that transcend context. For example, someone might believe that abortion is wrong, a belief that generalizes across time, context, and group boundaries. The distinction between what is normative versus imperative is not trivial: People readily perceive a psychological distinction between domains of normative standards and personal morality (Turiel, 1983), and each of these domains of social life are associated with unique perceived characteristics. Normative standards are perceived as culturally determined, as well as authority and situation dependent (e.g., the Orthodox Jewish belief that it is unclean for them to eat pork because religious authority and cultural norms say so, but that it is acceptable for other groups to eat pork given that they have different norms). In contrast, personal moral standards are perceived as universally applicable,

objectively true, and authority independent (e.g., the belief that female circumcision is objectively wrong, everywhere, regardless of what local norms or authorities deem right). Accordingly, people feel comfortable violating normative standards so long as relevant peers and authority figures condone it, but they firmly uphold their moral attitudes regardless of peers' or authorities' stance on the issue (Aramovich, Lytle, & Skitka, 2012; Skitka, Bauman, & Lytle, 2009; Turiel, 1983).

Still unclear is whether people sometimes perceive extreme acts of moral disengagement—including harming others—as flexible normative standards that can be violated under particular circumstances. Taking my theoretical position to the extreme, people's tendency to excuse norm violations when they serve a higher moral purpose could ultimately lead them to feel licensed to harm others. Indeed, what is "normative" and what is "morally mandated" is fluid and in the eye of the perceiver (Skitka, Bauman, & Sargis, 2005). Harming others can be perceived as a shockingly mundane act under some circumstances: specifically, when one thinks that doing so serves a higher order good. Parents, for example, seldom have difficulty exposing their children to the pain of vaccinations because they believe that the vaccination will protect their children from subsequent and more significant harm. The idea that norms that prohibit harming others can be overridden by a higher moral purpose has been echoed throughout history. A scholar of Nazi Germany, for example, noted, "The Final Solution did not develop as evil incarnate but rather as the dark side of ethnic righteousness. Conscience, originally seen to protect the integrity of the individual from the inhumane demands of the group, in the Third Reich became a means of underwriting the attack by the strong against the weak. To Germans caught up in a simulacrum of high moral purpose, purification of racial aliens became a difficult but necessary duty" (Koonz, 2003, p. 273). For Germans in the Third Reich who shared this

moral vision, mass genocide of "racial aliens" appears to have been perceived as a justifiable means toward achieving a higher moral end (purifying the German race), even if this norm violation is very likely to be perceived as unjustifiable and grossly immoral to modern readers who do not share the Nazi cause. In short, it appears that even extreme acts of moral disengagement can be construed as optional normative standards when moralized commitments or beliefs are at stake.

Morally Motivated Transgressions: Links to Moral Conviction

Although domains of normative standards and personal morality often coexist without conflict, there are undoubtedly instances when the two domains clash. Do these moments—that is, when people feel they must undermine normative standards to serve a personal moral conviction (e.g., when Americans feel they must harm enemies who threaten their moral conviction to live freely) —capture the essence of Bandura's (1999) construct of moral disengagement? Might the distinction between personal morality and normative standards clarify the apparent paradox of how people morally disengage to transgress for a moral cause with which they are seemingly highly engaged? The Integrated Theory of Moral Conviction (see Skitka, 2010 for a review) is poised from a theoretical standpoint to tackle the paradox of how people who are maximally morally engaged with a specific moral cause may ironically disengage from normative standards of right and wrong to serve that belief.

Moral conviction refers to the recognition that a specific attitude or belief is imbued with moral fervor (Skitka et al., 2005). Although some theories take a top-down approach to studying the psychological content of morality (e.g., implicitly assuming that specific situations trigger concerns with harm and fairness, e.g., Kohlberg, 1975; Turiel, Hildebrandt, & Wainryb, 1991), the moral conviction program of research has taken a different approach by asking whether

people see a given issue in moral terms, exploring what leads to this recognition, and the consequences of seeing various things in a moral light. This approach treats "moral" and "morality" as subjective impressions people have that distinguish some feelings, beliefs, or judgments from others. A perception of moral relevance is something I argue people bring to a given situation, rather than necessarily being an inherent quality of a given situation itself.

Many theorists, in contrast, treat some political issues (e.g., abortion) and dilemmas (e.g., trolley problems—Edmonds, 2013) as if everyone were to agree on their inherent relevance to morality. For example, Milgram (1974) interpreted the results of his studies on destructive obedience this way: "When asked to carry out actions incompatible with fundamental standards of morality, relatively few people have the resources to resist authority." But Milgram did not ask whether his participants perceived their actions as a trade-off between personal moral standards and obedience to authority; he assumed they perceived their actions in that way (see also Doris, 1998). I argue that to really know whether people are willing to sacrifice their moral standards to obey an authority (as just one example), one first has to know whether they interpret the situation in moral terms. Consistent with this idea, the degree to which people perceive that they have a personal moral stake in a given issue—that is, whether they see their position on a given issue as a reflection of their personal moral convictions—is associated with a number of defining features and consequences that cannot be explained by other attitude-related characteristics (e.g., attitude extremity, importance, certainty, centrality, or strength of partisanship).

Unlike non-moral attitudes, attitudes high in moral conviction (or "moral mandates") are theoretically associated with perceived universality, perceived objectivity, autonomy from the dictates of authority, and strong emotions (Skitka et al., 2005). Together, these elements are

thought to give moral mandates strong motivational force; moral mandates in and of themselves are believed to motivate individuals to take attitudinally-consistent action, even actions that are disengaged from *normative* standards (see Skitka, 2010 for a review). I elaborate on these features below.

First, moral mandates are perceived as more universal than other attitudes like preferences or normative conventions (Skitka et al., 2005). Unlike moral mandates, preferences are at the discretion of the individual and are not socially regulated. Also in contrast to moral mandates, normative conventions dictate how individuals within a certain social group are supposed to act; they do not speak to how *all* people should act (Turiel, 1983). Moral mandates are theoretically different from preferences and conventions because they are perceived to be absolute standards of right and wrong that are not culturally bound. Although people with moral conviction for an issue may acknowledge that others have different moral viewpoints, they believe that others would adopt such an absolute standard if they knew the "facts" (Skitka et al., 2005). The universalism hypothesis has received empirical support: People are more likely to have a universalistic mindset when thinking about their moral convictions and are also more likely to perceive their moral mandates as more universally applicable than their strong non-moral attitudes (Morgan, Skitka, & Lytle, under review).

Second, moral mandates are experienced as objective facts about the world, that is, people are likely to state that their moral mandates are objectively right or wrong without further justification. Moral mandates are perceived to be as factual as 2 + 2 = 4 (Skitka et al., 2005). The objectivity that is thought to be associated with moral mandates has also received empirical support. The more strongly an attitude is imbued with moral conviction, the more strongly one perceives that attitude to be objectively true (Morgan et al., under review).

Third, morally convicted attitudes have the defining characteristic of motivating people to act in accordance with these beliefs. Unlike non-moral preferences and conventions, the anticipated negative consequences of falling short of one's moral mandates are hypothesized to be quite strong, including feelings of shame, regret, and guilt. Similarly, adherence to moral mandates is thought to be associated with stronger positive outcomes (e.g., pride, gratification, elevation, and self-affirmation) than being true to one's preferences or following normative conventions. Research has demonstrated that strong emotions provide moral mandates the motivational force necessary to compel people to take action to support the attitudes: For example, positive affect partially mediated the relationship between moral conviction and activist intentions for supporters, whereas negative affect partially mediated the relationship between moral conviction and activist intentions for opposers of physician-assisted suicide (Skitka & Wisneski, 2011).

Ultimately, the motivational force associated with moral mandates could play a role in producing behavior that is disengaged from *normative* standards, including the decision to harm others who stand in opposition to the perceiver's moralized agenda. Most relevant is the finding that people with strong moral conviction for an issue are more intolerant of attitudinally dissimilar others than people who hold a strong but non-moral attitude toward the issue. People with strong moral conviction prefer greater social and physical distance from attitudinally dissimilar others, and they are less cooperative and agreeable in attitudinally heterogeneous group settings when moral mandates are at stake (Skitka et al., 2005). Their intolerance for the perspectives of others who threaten those moral mandates also make the morally convicted unwilling to compromise (Ginges, Atran, Medin, & Shikaki, 2007; Ryan, 2017; Skitka et al., 2005; Tetlock, Kristel, Elson, Green, & Lerner, 2000). Furthermore, people are willing to

behaviorally discriminate against others who do not share their moral beliefs: When asked to divide 10 raffle tickets between themselves and another attitudinally dissimilar person, people who were high (vs. low) in moral conviction were much more likely to keep most of the tickets for themselves (on average 8.5 tickets) rather than divide them equally (Wright, Cullum, & Schwab, 2008). In short, people with strong moral conviction are intolerant of attitudinally dissimilar others.

People with strong moral conviction for a cause are even willing to accept violence if it serves that higher moral purpose. For example, in the context of a capital punishment trial, participants' moral conviction about defendant guilt or innocence (i.e., that guilty murderers should be punished and the innocent be freed) had a greater impact on perceptions of the trial's procedural and outcome fairness than the actual procedures used. In short, people's strong moral convictions did a better job predicting perceptions of the trial's fairness than whether due process was upheld. Even when the procedures were shockingly unfair and violent (i.e., the defendant was killed by vigilantes), people's perceptions of the procedural and outcome fairness of what happened were still more strongly shaped by their morally vested beliefs about defendant guilt or innocence than they were by whether the procedures used to achieve them were procedurally fair or unfair. Due process was only rated as fairer than vigilante justice among people who had weak moral conviction for defendant guilt or innocence (Skitka & Houston, 2001; see also Bauman & Skitka, 2009; Skitka, 2002; Skitka et al., 2005; Skitka & Mullen, 2002).

Summary. Strong moral convictions embolden people to adhere strongly to their beliefs, even if that means they must be intolerant of attitudinally dissimilar others or accept violent means to achieve a morally convicted end. Although not often framed as such, one could conceptualize these findings as the *disengagement from normative standards* when people are

maximally engaged with specific moral convictions. In other words, people are ironically most likely to "morally disengage" (normatively) when they are maximally morally engaged (personally). Therefore, in my view, it is possible and instructive to interpret moral disengagement through the lens of moral conviction. People may feel compelled to take a stand for their strong moral convictions via normatively transgressive means: a behavior that could ironically affirm their sense of moral goodness.

Strong Moral Conviction Leads to Disengagement from Normative Standards

How can Bandura's (1999) theory of moral disengagement be interpreted through the lens of moral conviction? As reviewed above, the morally convicted are skilled at justifying any means necessary to uphold those beliefs; they perceive those means as morally imperative, even if the means themselves are perceived by others as immoral. Therefore, normative disengagement via maximal engagement with moral convictions seems to operate through the reconstrual of immoral behavior. People feel "off the hook" for their normative transgressions when they reconstrue those actions to be a negligible price to pay for a large moral payoff. In fact, Bandura (1999) theorized that people feel justified in transgressing because they truly believe they are upholding and protecting their cherished values.

Lost in the original nomenclature of "moral disengagement," however, is the critical idea that such a process nonetheless requires maximal engagement with issue-specific moral convictions. Indeed, people do not morally disengage because they decide to be "bad" and nefarious, or because they no longer care about being "good." People have a strong need to see themselves as moral (Bandura, 1989; Monin & Jordan, 2009; Steele, 1988, 1999), something likely to still be true in situations that appear to be examples of moral disengagement. Instead, people are sometimes forced to ignore normative standards when those standards collide with

their cherished moral convictions. In such circumstances, people may feel entitled to shirk

normative standards to serve a perceived higher moral purpose: their moral convictions. Thus,

even if they are disengaged from normative standards, people can still preserve their moral self-

integrity and moral self-concept by being engaged with their moral convictions.

Although not explicitly tested in the context of moral disengagement, there is existing

empirical evidence from the moral conviction program of research that supports my re-

conceptualization of moral disengagement. As explained above, to serve their moral convictions,

people are willing to: (a) deny the legitimacy of formal laws and authority (Bauman & Skitka,

2009; Mullen & Skitka, 2006; Skitka & Houston, 2001; Skitka & Mullen, 2002); (b) violate

social norms (Aramovich et al., 2012; Hornsey et al., 2003, 2007); (c) excuse the lies of corrupt

political figures (Mueller & Skitka, 2017); and (d) approve of violence to achieve morally

preferred ends (Reifen Tagar, Morgan, Skitka, & Halperin, 2013; Skitka & Houston, 2001).

Thus, it is clear that people are willing to violate some normative standards to serve their moral

convictions, even going so far as to condone harming others for the cause (see also Fiske & Rai,

2014).

Morally Motivated Transgressions: A Boost for Moral Self-Concepts

Many theories of moral motivation presuppose that a strong moral self-concept promotes

normatively upstanding behavior (Aquino & Reed, 2002; Blasi, 1983, 2004; Colby & Damon,

1992; Crocker & Wolfe, 2001; Hart, Yates, Fegley, & Wilson, 1995; Kraut, 1973; Lapsley &

Lasky, 2001; Reed & Aquino, 2003; Reed, Aquino, & Levy, 2007; Walker & Frimer, 2007;

Walker & Hennig, 2004): *not* normatively transgressive behavior like harming others who

disagree. Consistent with this idea, when people engage in normatively transgressive behavior,

predominant theories of moral behavior suggest that the transgressions should represent a threat

to their moral self-concepts (e.g., Monin & Jordan, 2009; Shao, Aquino, & Freeman, 2008). In contrast to these dominant theories of moral motivation, my dissertation studies will explore the possibility that normatively transgressing in the name of a strong personal conviction may not negatively impact people's sense of moral integrity. In fact, because people take a risk to affirm their moral convictions—that is, violating normative conventions—they may ultimately construe morally motivated transgressions to be *morally courageous* and a boost to their moral self-concept.

The possibility that people construe their morally motivated transgressions as morally courageous is consistent with dominant definitions of moral courage. Traditionally, moral courage has been defined as taking a stand for a principle or conviction, even if taking a stand means exposing oneself to potential risks such as "inconvenience, unpopularity, ostracism, disapproval, derision, and even harm to [oneself] or [one's] kin" (Skitka, 2012, p. 4; see also Miller, 2000). Moral courage is, "…less about risks, hazards, obstacles, and [more] about values, virtues, standards, and rightness" (Miller, 2000, p. 36).[1] More specifically, moral courage situations are theorized to involve the following social dynamic: (1) one or more perceived perpetrators who violate a moral principle; (2) one or more perceived victims who suffer as a result of a violated moral principle; and (3) a socially risky confrontation between the morally courageous helper and the perpetrator(s). This third feature—helpers who expect negative social consequences for their intervention—is what distinguishes moral courage from other forms of

---

[1] Moral courage is similar to the construct of heroism, in that both constructs involve taking a stand at great personal risk (e.g., the possibility of getting hurt while defending an ethical principle). Where these two constructs diverge, however, is that people who act heroically expect widespread positive social consequences for their actions (e.g., admiration) whereas people who act morally courageous can expect negative social consequences (e.g., persecution; Osswald, Greitemeyer, Fischer, & Frey, 2010).

helping behavior, such as heroism (Greitemeyer, Fischer, Kastenmueller, & Frey, 2006; Greitemeyer, Osswald, Fischer, & Frey, 2007; Osswald et al., 2010). To the extent that morally motivated transgressions are perceived as taking a personal risk to stand up for a cause that is at stake, then transgressors may feel morally courageous for their behavior. In other words, these transgressions might ironically feel courageous to perpetrators.

That said, morally courageous behavior is often assumed by researchers to be inherently prosocial. For example, moral courage is considered by some to be a branch of prosocial/helping behavior (e.g., Niesta Kayser, Greitemeyer, Fischer, & Frey, 2010), and researchers of moral courage often study moral exemplars who are deemed as exceptionally virtuous by society as a whole (e.g., national awardees for exceptional bravery, Walker & Frimer, 2007). These perspectives on moral courage tend to take a top-down approach, implicitly defining the classes of behaviors that are morally courageous. In doing so, however, they ignore the possibility that actors engaging in morally motivated transgressions could appraise their own behavior as morally courageous as well. I propose that it is possible to feel morally courageous by serving a strong moral conviction at great personal risk: even if that means that perceivers who disagree may not appreciate that moral conviction or the advocacy used to bolster the cause.

Following risky displays of moral courage—for example, when people defend a moral conviction via normatively transgressive means—it is plausible that people's moral self-concept may be temporarily inflated. Indeed, moral self-concepts are dynamic and ever-changing depending on people's actual behavioral history or construals of recent behavior (Monin & Jordan, 2009). People can maintain a positive moral self-concept by actually being a normatively good person, or by construing their behavior in a biased way: as more morally upstanding than it actually is (e.g., Monin & Miller, 2001). To the extent that people construe their morally

motivated transgressions to be in service of the greater good at personal risk, they may: (1) downplay the transgressiveness of their behavior, (2) focus on the morally mandated outcome achieved, (3) ultimately construe the behavior to be morally upstanding, and (4) temporarily enjoy an inflated moral self-concept. This psychological outcome of morally motivated transgressions would suggest a dark side to moral identity and maximal moral engagement: that people can convince themselves that the means themselves matter less than achieving a morally convicted outcome, even if the cause is deemed to be misguided by observers.

Non-Moralized Transgressions: Alternative Mechanisms

As described above, stronger moral conviction for a cause that is at stake should predict greater motivation to serve the cause. When this strong motivation to serve a moralized cause is in conflict with normative standards, I predict that people should be willing to violate those norms, construe their transgressions in a positively biased way (i.e., as supportive of a higher moral purpose), and enjoy a boost in their moral self-concepts. But what happens when people feel strongly about serving a cause that they *do not* imbue with strong moral conviction?

Bandura (1999) theorized that there are two non-moral paths that license reprehensible behavior. First, people may feel comfortable transgressing for a cause because they use sanitizing language or euphemisms to soften the blow of their actions. Atrocities that are cloaked in language that hides their true repugnancy (e.g., saying that soldiers "waste" wartime enemies instead of using the more accurate terms "kill" or "murder") are thought to be easier to commit than atrocities without euphemistic labeling (Bandura, 1999). Supporting this notion, people are more vicious when aggressions are framed in euphemistic (vs. accurate) language (Diener, Dineen, Endresen, Beaman, & Fraser, 1975): for example, when people's physical aggression is framed as being part of a "game," they are crueler than when their actions are labeled

"aggressive" (Diener et al., 1975).[2] Second, people may feel like they are entitled to engage in reprehensible conduct when they compare their own behavior to that of an even greater villain. By contrasting one's own actions against another person or group's extreme atrocities, transgressions are theorized to be perceived as negligible or even upstanding (Bandura, 1999).

It is unclear, however, the extent to which non-moral paths to reprehensible conduct can operate without the co-occurrence of moral justification. I contend that non-moral paths alone do not license people's engagement in reprehensible conduct, for two reasons. First, people with strong but non-moral attitudes for an issue do not have the same level of motivation to transgress for that cause as those who hold strong moral attitudes. For example, people with no moral investment in the cause at hand tend to cooperate with rather than be intolerant of attitudinally dissimilar others with whom they are tasked to work with (Skitka et al., 2005), are less discriminating against others who disagree (Wright et al., 2008), and avoid supporting immoral, hostile forms of collective action (Zaal, Van Laar, Stahl, Ellemers, & Derks, 2011), whereas people with a high moral investment are disinclined to do any of these things. Together, these findings suggest that strong but non-moral attitudes do not provide compelling motive to break normative standards (e.g., group (un)cooperation, inequitable division of resources, or most pertinent to my dissertation studies, harming others).

Second, harming others via non-moral paths alone should represent a substantial threat to people's moral self-concepts, something that should deter them from transgressing (e.g., Monin & Jordan, 2009; Shao, Aquino, & Freeman, 2008). For example, in the absence of genuine moral

---

[2] It is important to note that this study was primarily aimed to investigate the deindividuation of individuals within a group when euphemistic labeling of the aggressive task was present or absent. To the best of my knowledge, it is unclear whether euphemistic labeling licenses *individuals'* harmful acts independent of the prevailing group dynamics present in Diener et al. (1975).

justification, being provided a euphemistic label for a harmful act should not likely lead people to reconstrue the act as prosocial and something reflective of strong moral integrity; people are only able to deceive themselves of the righteousness of their acts when the acts themselves are perceived as morally ambiguous, not unambiguously harmful (Miller & Effron, 2010). Therefore, harmful acts that are euphemized should still likely be threatening to people's moral self-concepts when they have no moral investment in the cause. Similarly, although downward social comparisons generally boost people's subjective sense of well-being (Wills, 1981), people with little moral motive to transgress may not be likely to accept harmful villains as appropriate references for social comparison and may therefore fail to subsequently expect a boost in moral self-regard: People tend to evaluate themselves in reference to similar others to get an accurate sense of their own standing (Festinger, 1954; Goethals & Darley, 1977; Wheeler, Martin, & Suls, 1997; Smith & Zarate, 1992).

Advantageous downward comparisons and euphemistic labeling may instead only further justify reprehensible conduct when paired with moral justification. After people disengage from normative standards to serve their strong moral convictions, euphemistic labeling and advantageous comparisons could further bolster moral self-regard by: (a) reinforcing people's reconstrual of the transgressiveness of their actions, and (b) providing an appropriate target for downward moral comparison, which should further increase moral self-esteem (Wills, 1981).

Taken together, I theorize that moral investment in a cause is the primary path toward reprehensible behavior that provides inherent motivation for belief-bolstering transgressions while protecting people's sense of moral integrity. People should be most willing to engage in reprehensible behavior for a cause when they have a moral stake in it, lest they severely threaten their own sense of moral goodness: something that people are extremely motivated to avoid

(Bandura, 1989; Monin & Jordan, 2009; Steele, 1988, 1999). In short, moral justification should be the primary path toward licensing reprehensible behavior, such as harming others.

Exploratory Questions

Bandura (1999) theorized that several other mechanisms shape the psychological *aftermath* of deciding whether to engage in reprehensible behavior, including people's perceptions of the consequences of their actions and the victims they harmed (see Figure 1). An exploratory aim of my dissertation will be to examine the extent to which these factors influence people's *prospective* willingness to engage in morally motivated transgressions.

Specifically, after engaging in reprehensible behavior, people are theorized to ignore, minimize, distort, or disbelieve evidence that what they did was harmful. This seems to be easier to do when people do not see the consequences of their actions firsthand (Bandura, 2011; Milgram, 1974; Royakkers & Van Est, 2010; Tilker, 1970). Moreover, perhaps the easiest way to deny the harmful consequences of one's actions is to dehumanize victims: to strip them of their human qualities (i.e., personal identity and connection to others) and to cast them as subhuman. Perceiving one's enemies as subhuman facilitates the denial that one's actions are harmful and morally reprehensible, and propels people to more strongly support retaliatory action in the context of real intergroup violence (e.g., the Boston Marathon bombings; Kteily, Bruneau, Waytz, & Cotterill, 2015).

Still unclear is the extent to which these mechanisms not only shape the psychological aftermath of reprehensible conduct, but also people's decision to commit harmful acts in the first place. Indeed, knowing that one will have expedient opportunity to deny harmful consequences (e.g., by harming enemies remotely) may make it easier to *decide* to harm others. Similarly, the perception that potential victims are subhuman likely makes it easier to *decide* to transgress

against them (Kelman, 1987; Opotow, 1990; Rai, Valdesolo, & Graham, 2017; Struch & Schwartz, 1989). In short, although Bandura (1999) assigned a temporal order to these mechanisms of moral disengagement, my dissertation will explore the extent to which all of them can provide prospective justification for deciding to engage in morally motivated transgressions in concert once people have a moral investment in the issue at hand (see Figure 2; cf. Rai et al., 2017).

Before turning to my hypotheses, it is important to note that my dissertation focuses on harms intended to achieve political ends and the processes that drive endorsement of such harms. My theory does not account for harms that are functional for survival, such as self-protection/self-defense. I do not anticipate that my theorized processes contribute to endorsement of *all* harms, including self-defense. Rather, I believe that the processes outlined above uniquely apply to situations in which actors are politically motivated and are attempting to harm others to achieve a politically motivated agenda.
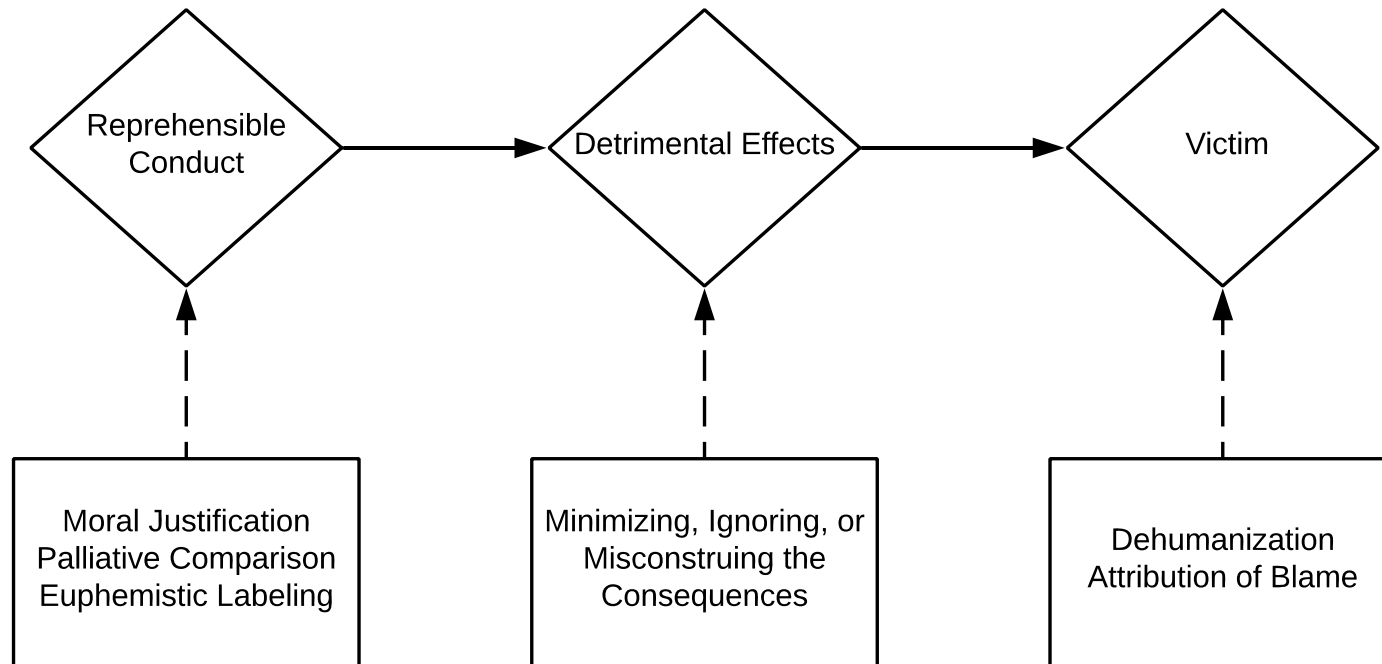
Figure 1. Bandura's (1999) theory of moral disengagement (adapted from Bandura, 1999).
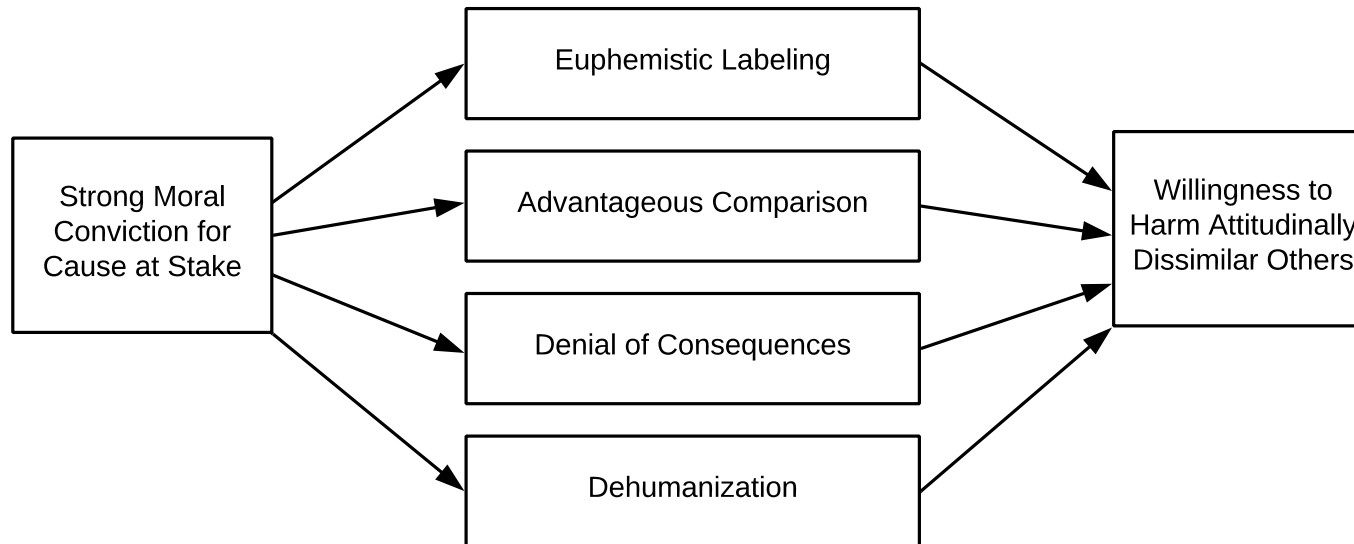
Figure 2. The Moral Primacy Hypothesis. Strong moral conviction for a cause at stake is the primary factor driving people's willingness to harm attitudinally dissimilar others. Bandura's other (1999) theorized disengagement processes should only play secondary roles as mediators in the link between strong moral conviction and willingness to endorse harmful actions against political opponents.

Primary Hypotheses

Taken together, the *moral motivation hypothesis* predicts that people will be more willing to harm others in an effort to achieve strongly (vs. weakly) moralized ends, a process reflective of Bandura's (1999) moral justification pathway to reprehensible conduct. Given that harming others for a strong (vs. weak) moral conviction is more likely to be perceived as a risky display of commitment for the cause, people should be more likely to construe their transgressiveness as a morally courageous gesture and as a reflection of their strong moral integrity. The ultimate consequence of this moral pathway, then, is a boost in people's moral self-regard when people imbue the cause at stake with strong (vs. weak) moral conviction (see Figure 3).

Moreover, in contrast to Bandura's (1999) assertion that there are independent non-moral pathways toward reprehensible conduct, the *moral primacy hypothesis* predicts that moral investment in a cause is the primary path to reprehensible behavior that provides inherent justification for belief-bolstering transgressions while protecting people's sense of moral integrity. Other non-moral justifications, such as euphemistic labeling and advantageous comparisons (as well as exploratory pathways such as denial of consequences and dehumanization of enemies), should only be invoked when people already have strong moral justification that their reprehensible action will serve a higher order good. In other words, the *moral primacy hypothesis* predicts that non-moral justifications will only predict harm endorsement when people already have a strong (but not weak) moral investment in the cause.

Finally, a corollary of the *moral primacy hypothesis* is that non-moral justifications for norm disengagement, including denial of consequences and dehumanization of victims, will mediate the link between people's strong moral conviction for the cause at stake and their prospective willingness to harm others who stand in opposition to that cause. In other words, the

moral primacy hypothesis presents an alternative process model from Bandura's (1999) theory of

moral disengagement, whereby all non-moral justifications should operate in concert to further

justify harming attitudinal opponents once people already have a strong (but not weak) moral

justification for harmful action (see Figures 2 & 4).
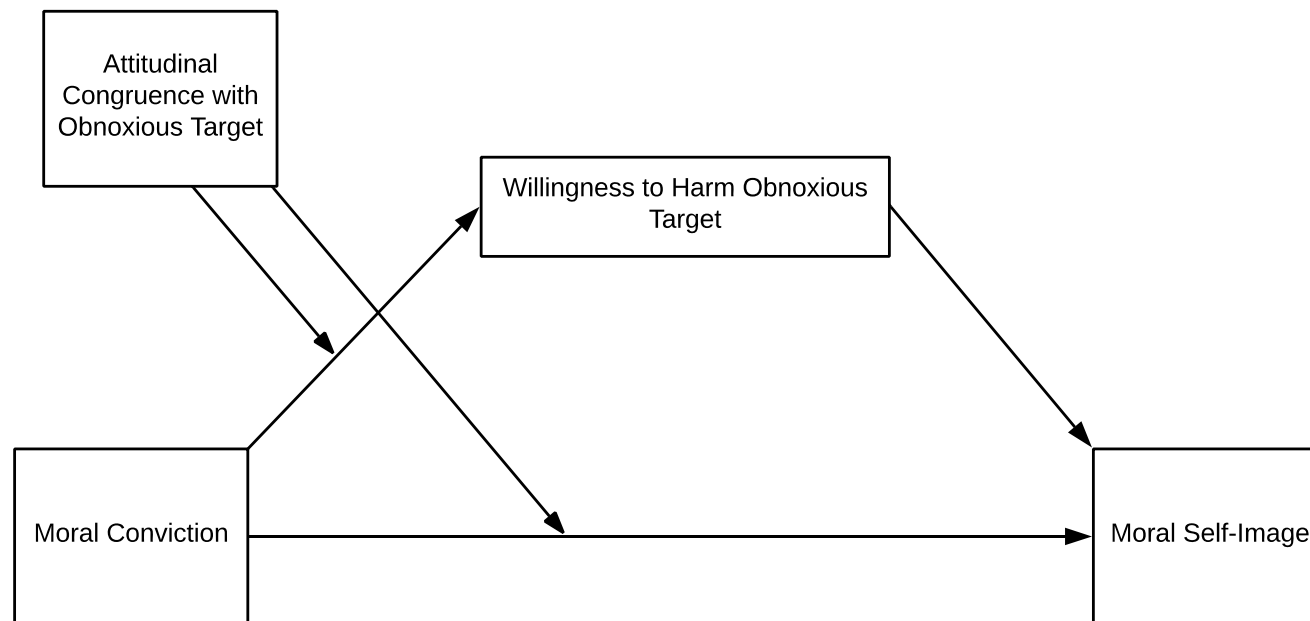
Figure 3. The Moral Motivation and Incongruence Hypotheses: The predicted mediation pattern between moral conviction, willingness to harm obnoxious targets, and moral self-image, moderated by participants' attitudinal congruence with the obnoxious target. The direct effect between moral conviction and moral self-image should emerge only when participants *oppose* (but not support) the attitude of the target. Moreover, the indirect link between moral conviction and moral self-image through willingness to harm should similarly only emerge when participants *oppose* (but not support) the target's attitude.
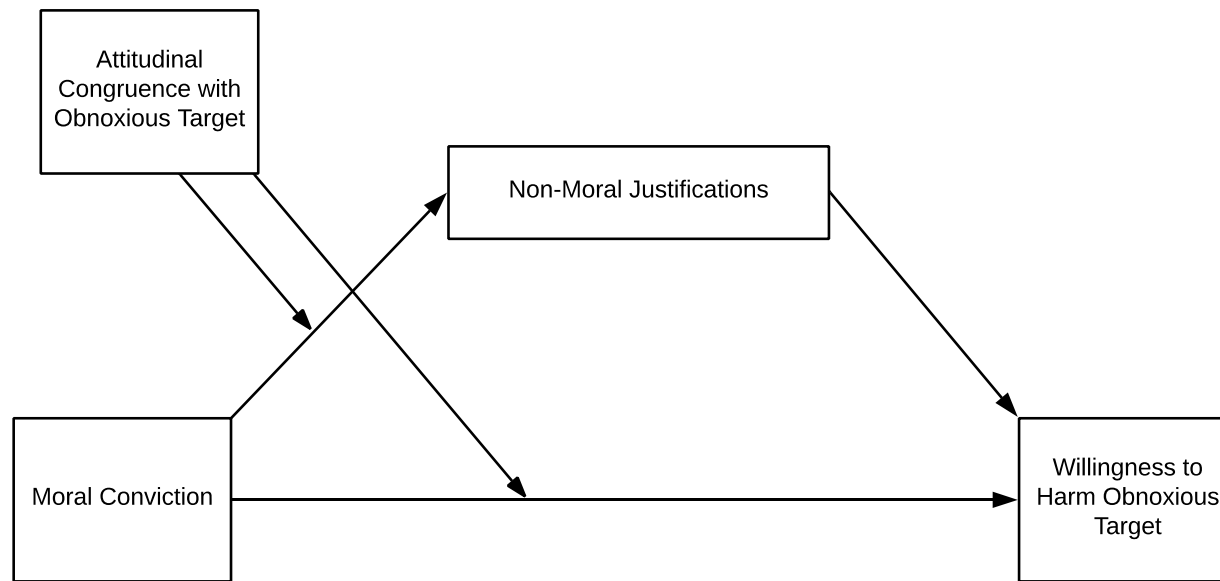
Figure 4. Incongruence & Ideological Symmetry Hypotheses: The predicted mediation pattern between moral conviction, non-moral justifications, and willingness to harm obnoxious targets, moderated by participants' attitudinal congruence with the obnoxious target. The direct effect between moral conviction and willingness to harm obnoxious targets should emerge only when participants *oppose* (but not support) the attitude of the target. Moreover, the indirect link between moral conviction and willingness to harm through non-moral justifications should similarly only emerge when participants *oppose* (but not support) the target's attitude. I predict that this moderated mediation effect should be symmetrical for *both* liberals and conservatives/conservative and liberal targets.

Secondary Hypotheses

Are People Willing to Harm Obnoxious Members of their Attitudinal Ingroup?

Thus far in my review, I have described how strong moral conviction for a cause may embolden people to harm others who stand in the way of that value, that is, *attitudinal opponents*. However, the Black Sheep Effect suggests that people derogate members of their ingroup who threaten the overall reputation of the group (Marques, Yzerbyt, & Leyens, 1988). Are people willing to harmfully derogate members of their own side who go too far in supporting a shared cause? I theorize that people will *not* be willing to harm members of their attitudinal ingroup: People applaud transgressive ingroup members who go too far to bolster a shared cause, especially in politicized contexts when the cause is imbued with strong moral conviction (Mueller & Skitka, 2017; cf. Mullen & Skitka, 2006). I predict that the psychological levers that may license harmful action toward attitudinal opponents *should not* license harms against attitudinal ingroup members. In other words, attitudinal congruence/incongruence with a target should be an important moderator of my predicted model. Stronger moral conviction should predict greater willingness to harm obnoxious attitudinal outgroup members, but not attitudinal ingroup members. This hypothesis will be referred to as the *incongruence hypothesis*.

Are Liberals and Conservatives Willing to Harm Opponents to the Same Degree?

Conservatives have historically been characterized by social psychological researchers as more prejudiced and intolerant than liberals (Sibley & Duckitt, 2008), due to their close-mindedness (e.g., Jost, Glaser, Kruglanski, & Sulloway, 2003). Based on this reasoning, one could be tempted to argue that conservatives are more likely to harm their attitudinal opponents than liberals when a strong moral conviction is at stake.

However, a growing stream of alternative findings suggests that liberals and conservatives may be equally intolerant toward targets whose values are inconsistent with their own (i.e., the ideological conflict hypothesis, Brandt, Reyna, Chambers, Crawford, & Wetherell, 2014). According to this perspective, conservatives' historic intolerance is merely an artifact of biased methodologies. These methodologies tended to ask people to evaluate intolerance toward traditionally liberal groups who threatened conservative values, but not conservative targets who threatened liberal values (Brandt et al., 2014). More recent research utilizing a variety of target groups (both liberal and conservative) find support for the ideological conflict hypothesis. Conservatives (vs. liberals) tend to express more intolerance toward liberal targets (e.g., prochoice advocates and welfare recipients), whereas liberals (vs. conservatives) tend to express more intolerance toward conservative targets (e.g., prolife advocates and Tea Party supporters; Brandt et al., 2014; Chambers, Schlenker, & Collisson, 2013; Crawford & Pilanski, 2014; Wetherell, Brandt, & Reyna, 2013). Willingness to express intolerance is driven in part by people's perceptions that their values and worldviews are threatened, a process that occurs on *both* the political right and left (Crawford & Pilanski, 2014; Wetherell et al., 2013).

Taken together, as researchers have expanded the scope of target groups beyond those that are traditionally liberal, it has become increasingly apparent that willingness to express intolerance toward those who violate important worldviews is symmetrical across the political divide (Brandt et al., 2014). I contend that ideological symmetry in intolerance should also extend to situations involving harm toward worldview-threatening attitudinal opponents. If true, then I should find that willingness to harm attitudinal opponents occurs to the same degree for political conservatives and liberals.

Moreover, I predict that the primacy of moral justification in driving people's willingness to harm attitudinal opponents should be equivalent for liberals and conservatives. Liberals and conservatives are equally likely to view political issues through the lens of strong moral conviction, and moral conviction is an equal opportunity motivator of political engagement (Skitka, Morgan, & Wisneski, 2015). I argue that viewing political issues through the lens of morality should be equivalent for liberals and conservatives, which should ultimately produce a similar tendency to engage in dark forms of political engagement (i.e., harming opponents to serve a higher moral purpose). This prediction will be referred to as the *ideological symmetry hypothesis*.

Summary of incongruence & ideological symmetry hypotheses. I predict that, when participants' attitudes are incongruent (vs. congruent) with the targets' (i.e., when participants *oppose* but not support the targets' attitudes), moral conviction for that issue should predict non-moral justifications for harm, which in turn should predict willingness to harm them: a pattern that should emerge for liberal *and* conservative participants (see Figure 4).

Do Perceptions of Threat Exacerbate People's Willingness to Harm Political Opponents?

The more that information, events, or experiences seem to undermine important values, beliefs, and identities, the more that people experience them as threatening (Proulx & Heine, 2010). In response to these meaning/worldview threats, people attempt to manage their anxiety. One way to do this is by derogating outgroups with dissimilar values and beliefs (Crawford, 2014; van Prooijen, Krouwel, Boiten, & Eendebak, 2015), a pattern that is similar among liberals and conservatives (see Crawford, 2017 for a review).

Based on this reasoning, I expect that exposure to divergent political attitudes (e.g., at a contentious political rally) represents a potential threat to people's worldviews. The more

strongly that people perceive such events as threatening to their worldviews, the more likely they

should be to engage in defensive compensatory reactions designed to manage feelings of anxiety.

One such reaction is by derogating political opponents, for example by expressing greater

willingness to harm political opponents than people not in a highly threatened state. Taken

together, I anticipate that perceptions of threat should moderate the link between strong moral

conviction for a cause and willingness to harm attitudinal opponents: The link should become

stronger when the attitudinal opponents are perceived as highly threatening (vs. lower levels of

perceived threat; see Figure 5). This prediction will be referred to as the *threat hypothesis.*
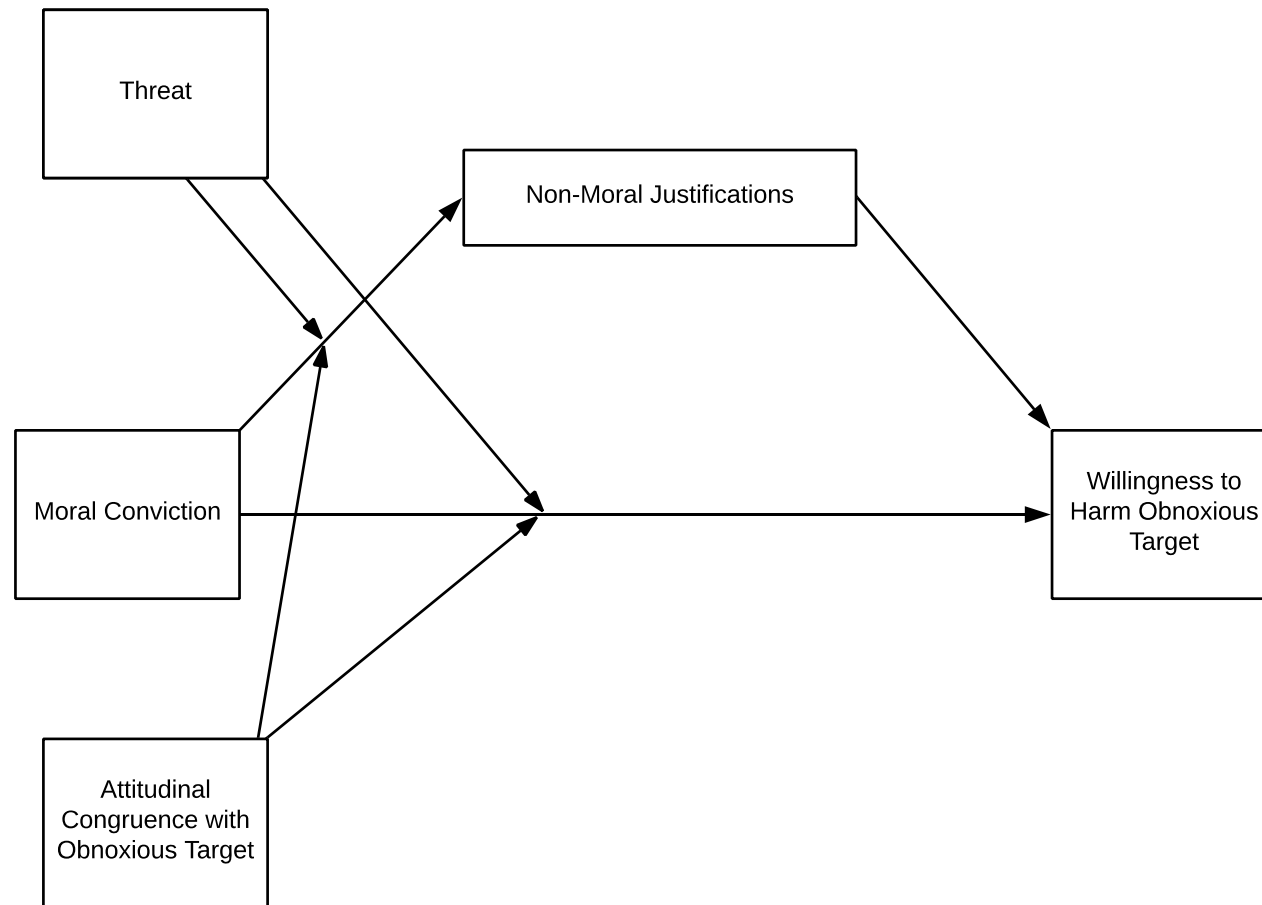
Figure 5. Threat Hypothesis: The predicted mediation pattern between moral conviction, non-moral justifications, and willingness to harm others, moderated by participants' perceived levels of worldview threat. The direct effect between moral conviction and willingness to harm oppositional (vs. likeminded) obnoxious targets should become stronger at higher levels of perceived threat. Moreover, the indirect link between moral conviction and willingness to harm opponents through non-moral justifications should similarly become stronger at higher levels of perceived threat.

Method Overview

To evaluate the factors that shape people's willingness to harm others for a cause, I asked participants to personally endorse varying levels of harm aimed against political targets at an Alt-Right rally/protest and explored factors that predicted those judgments. Specifically, I presented participants a conservative target in Study 1 (obnoxious opponents of multiculturalism), and a liberal target in Study 2 (obnoxious supporters of multiculturalism). Before I used this approach to test hypotheses, however, it was important to pilot test a corpus of harmful behaviors that people were personally willing to endorse for a cause. Pilot testing how far people were willing to go to aggress against protesters ensured that the stimuli used in Studies 1 and 2 were believable, tempting to endorse, and something that would enhance variance in responding.

Pilot Study

As described above, the first goal of the Pilot Study was to uncover a set of behaviors characteristic of those observed at contentious political rallies (i.e., Alt-Right/Antifa rallies) that varied in levels of harm inflicted toward others. Overall, the Pilot Study was necessary to identify a corpus of harmful behaviors that participants were tempted to personally endorse in Studies 1 and 2. A second aim of the Pilot Study was to confirm that the scales I administered in the Main Studies—including justifications for harm and perceptions of worldview threat at the rallies—were reliable. A third aim of the Pilot was to explore the degree to which a specific political issue—multiculturalism—was perceived as a core issue at stake at Alt-Right/Antifa rallies, and the degree to which support/opposition to the issue was associated with liberalism/conservatism. It was necessary to confirm that multiculturalism was perceived as an issue at stake at the rallies so that moral conviction for multiculturalism would be an appropriate attitude to measure in my Main Studies/a potential motivator of violence toward opponents at the

rallies. Moreover, in the Main Studies, I tested the ideological symmetry hypothesis: that people with liberal/conservative stances would be willing to harm oppositional conservative/liberal targets to equivalent degrees. Given that participants were evaluating supporters and opponents of multiculturalism, it was necessary to confirm that *support* for multiculturalism was spontaneously associated with *liberals*, and *opposition* associated with *conservatives*; targets who supported multiculturalism should have been perceived as liberals, whereas targets who opposed multiculturalism should have been perceived as conservatives.

To accomplish all of these goals while reducing individual participant burden and fatigue, I broke the pilot into two shorter surveys with separate groups of participants. The first survey primarily assessed the extent to which participants perceived Appendix A behaviors to be harmful. The second survey primarily assessed the extent to which participants endorsed Appendix A behaviors and their justifications for endorsement. I elaborate on these procedural differences between the two versions of the Pilot Study below. Lastly, I pooled both samples together to: (a) assess perceptions of a past Alt-Right rally video that would be shown to Main Study participants to increase their engagement; (b) assess multiculturalism as an attitude object for the Main Study and its relevance to Alt-Right/Antifa rallies; and (c) assess the extent to which support for/ opposition to multiculturalism is perceived as a liberal/conservative position.

Participants

For the harmfulness version of the pilot study, I recruited 110 workers from Amazon's Mechanical Turk (U.S. only; HIT approval rate > 90%). In return for their participation, participants were paid $0.10/minute ($1.20 total). On average, the sample skewed slightly liberal ($M = -1.16$, $SD = 2.71$), on a -4 (*very much liberal*) to +4 (*very much conservative*) scale. See Table 1 below for the distribution of participants' political orientation.

For the endorsement/justifications version of the pilot study, I recruited 224 new workers from Amazon's Mechanical Turk (U.S. only; HIT approval rate > 90%). In return for their participation, participants were paid $0.10/minute ($1.20 total). On average, the sample was politically neutral ($M = 0.10$, $SD = 2.68$), on a -4 (*very much liberal*) to +4 (*very much conservative*) scale. See Table 2 below for the distribution of participants' political orientation.

Table 1
Harmfulness Pilot: Participants' Political Orientation

Political Orientation	Frequency	%
Very much liberal	25	23.6
Much liberal	19	17.9
Moderately liberal	17	16.0
Slightly/lean liberal	10	9.4
Neutral/neither	7	6.6
Slightly/lean conservative	8	7.5
Moderately conservative	3	2.8
Much conservative	4	3.8
Very much conservative	13	12.3

Table 2
Endorsement/Justifications Pilot: Participants' Political Orientation

Political Orientation	Frequency	%
Very much liberal	27	12.6
Much liberal	18	8.4
Moderately liberal	32	14.9
Slightly/lean liberal	20	9.3
Neutral/neither	12	5.6
Slightly/lean conservative	23	10.7
Moderately conservative	26	12.1
Much conservative	35	16.3
Very much conservative	22	10.2

Procedure

For both versions of the Pilot, participants entered the survey and completed the consent process, as well as a Captcha question (see Appendices B & C) to confirm that they were human participants. After participants consented to participate via Qualtrics, they were presented with a definition of multiculturalism: "Multiculturalism means that all different groups within society

are treated equally and have the same rights, and their cultural perspectives are given equal value

and status (e.g., Muslims/Christians, Blacks/Whites, gay and straight people, etc.).” They were

then asked, “Do you support or oppose multiculturalism in the U.S.?” with the options *support*,

oppose, or *neutral/neither*. If participants selected *neutral/neither*, they saw the follow-up

question, “If you had to say which way you lean, would you say you support or oppose

multiculturalism in the U.S.?” Response options for this question included *lean toward support*,

lean toward oppose, and *neutral/neither*. If participants initially selected *support* or *oppose*, they

were asked the follow up question: “To what extent do you support/oppose multiculturalism in

the U.S.?” with response choices *slightly*, *moderately*, *much*, and *very much*. Participants who

leaned toward supporting or opposing the issue were folded into the *slightly* support (or oppose)

scale point. Only participants who answered with some degree of support or opposition to the

issue were invited to complete the rest of the survey (something I also screened for in Studies 1

and 2 to ensure that I could classify participants as congruent or incongruent with the targets’

stance). In total, there were only two participants who had no attitude toward multiculturalism in

the harmfulness version of the Pilot, and only eight participants in the endorsement/justifications

version of the Pilot. I then asked participants to report their political orientation (see Measures

section for more detail).

   Those who reported that they supported or opposed the issue to some degree were

presented with a description of an upcoming Alt-Right rally that is slated to actually occur in

Charlottesville, Virginia (see: https://www.mercurynews.com/2018/07/24/alt-right-organizer-of-

charlottesville-march-drops-bid-for-second-rally/):

   “In 2019, political activists are gearing up for an Alt-Right rally that may be one of the

   most heated political rallies of our time. On one side are members of the Alt-Right, who

rally to protect the rights of White people in the United States by opposing

multiculturalism, immigration, and political correctness. Members of the other side,

including the left-wing anti-fascist group known as Antifa, rally to protect the rights of

minorities in the United States by supporting multiculturalism, lenient immigration laws,

and political correctness rules. The 2019 rally is currently being planned for

Charlottesville, Virginia. Largely considered to be the continuation of the Charlottesville

'Unite the Right' rally, the rally is perceived by some as a tactic to normalize white

nationalism. Members of Antifa—the left-wing anti-fascist group known to use force to

try to silence white supremacists—are expected to gear up for what may be a

reincarnation of the 2017 'Unite the Right' rally."

On the next screen, participants read: "Rallies with both protestors and counter-protestors lead to

conflict between members of opposing sides. Please watch the video below to see how conflict

swelled during the 2017 Unite the Right Rally in Charlottesville, Virginia." Participants then

watched a one-minute YouTube video that depicted on-the-ground conflict at the 2017 Unite the

Right Rally. This video was selected because it shows harmful behaviors displayed by *both* Alt-

Right and counter-protesters, including pushing, punching, and the use of pepper spray (see:

https://www.youtube.com/watch?v=26lkrCzObpQ&feature=youtu.be). Importantly, it does not

display extreme acts of violence, such as a car crashing into a crowd of protesters, so as not to

misconstrue typical behaviors at the rally as extremely violent.

Immediately following the video, participants answered questions about their impressions

of it, including an open-ended question about their reactions ("What are your impressions of the

video you just watched?"). They also answered close-ended questions about the extent to which

they believed the rally was violent: "To what extent do you think that the Unite the Right rally

depicted in the video was violent?" and the degree to which protesters within the video were

obnoxious: "To what extent do you think that the protesters' behavior in the video was

obnoxious?" Response options for these questions included *not at all*, *slightly*, *moderately*, *much*,

and *very much*.

For the next step of the procedure, the pilot studies diverged.

For the endorsement/justifications version of the Pilot, participants considered the various

behaviors they would be comfortable endorsing against protesters with *opposite* attitudes as

themselves. They read: "Imagine you attend the upcoming Alt-Right rally and counter-protests.

What kind of actions would you be comfortable taking against [pipe in opposing stance:

supporters/opponents of multiculturalism] at the rally who display behavior like that depicted in

the video? That is, to what degree do you feel this behavior is an appropriate reaction against

[supporters/opponents of multiculturalism] who display behavior like that depicted in the video?

For the remainder of today's study, you will be asked to consider your reactions to behaviors

between protesters." First, I asked participants to rate the extent to which they endorsed a variety

of harmful behaviors that could take place at political rallies (see Appendix A). For each

behavior, participants were asked, "To what extent can you imagine a set of circumstances where

an appropriate reaction for protesters like you would be to enact the following behaviors toward

[pipe in opposing ideology: supporters/opponents of multiculturalism] who display behavior like

that depicted in the video?" with response options *not at all*, *slightly*, *moderately*, *much*, and *very*

much. Second, because I administered a scale to assess the extent to which various

justifications influenced harm endorsement ratings in Studies 1 and 2, I used

this Pilot Study to confirm that it is reliable. Specifically, participants read: "Thinking about

your responses overall, how much do you agree that each factor below influenced

your perceptions of these behaviors?" They were then presented with a series of

statements inspired by Bandura's (1999) mechanisms. Moreover, embedded in these items was

a pretest of seven items designed to assess the degree to which participants perceived the rally

as threatening to their worldviews: a potential important moderator of my proposed model. See

Measures section below for more detail.

In the harmfulness version of the Pilot, new participants evaluated the behaviors in

Appendix A. They were asked to rate the extent to which they are harmful to their opponents.

They read: "To what extent are the following behaviors harmful to [pipe in opposing stance:

supporters/opponents of multiculturalism] who display behavior like that depicted in the video?"

Participants were presented with Appendix A behaviors, with the response options *not at all*,

slightly, *moderately*, *much*, and *very much*.

The final steps of the procedure were identical across both versions of the Pilot.

Finally, participants reported their impressions of the issue of multiculturalism in relation

to Alt-Right/Antifa rallies. I asked them the extent to which they saw the issue to be at stake at

Alt-Right/Antifa rallies, as well as the degree to which they associated support and opposition to

multiculturalism with liberals/conservatives. See Measures section below for more details.

During this last block of questions, participants also reported on the extent to which we should

trust their data (to ensure that they were paying attention). Participants were asked: "To what

extent should we trust your data?"; "How distracted were you while completing this study? (R)";

and "To what extent were you paying attention in this study?" Response options for all these

questions included *not at all*, *slightly*, *moderately*, *much*, and *very much*. As a final question to

make sure participants were paying attention, they were asked: "What is the topic of this study?

Sometimes participants do not carefully read the instructions. To correctly answer this question,

please select the option "other" and write down the name of your favorite movie." Response options included *sexism*, *multiculturalism*, *I don't remember*, and *other* with a text-entry field next to it. Participants passed this attention check if they wrote a movie title in the "other" field.

Measures

Political orientation. Participants' political orientation was assessed with: "Are your political beliefs generally liberal or conservative?" with the options *liberal*, *neutral/neither*, or *conservative*. If participants selected *neutral/neither*, they saw the follow-up question, "If you had to say which way you lean, would you say you are more liberal or conservative?" Response options for this question included *lean toward liberal*, *lean toward conservative*, and *neutral/neither*. To measure political extremity, among participants who initially selected *liberal* or *conservative*, they were then asked, "To what extent are your political beliefs liberal/conservative?" Response options for this question included *slightly*, *moderately*, *much*, and *very much*. Participants who leaned towards liberalism or conservatism were folded into the *slightly* liberal (or conservative) scale point.

Factors driving harm endorsement. To pilot a measure examining the extent to which each of Bandura's (1999) theorized mechanisms drive harm endorsement, I presented participants with the prompt, "Thinking about your responses overall, how much do you agree that each factor below influenced your perceptions of these behaviors?" They then saw 28 randomized statements (seven per theorized mechanism). For advantageous comparison they rated the extent to which the following statements influenced their judgments with the stem "To what extent do you think…" followed by the completions, "Other protesters would normally behave worse than this"; "Other people will do worse things than this at the upcoming rally"; "These behaviors are better than how protesters normally act at rallies"; "No one else I know

would behave in these ways (R)"; "These behaviors are worse than how people normally act at rallies (R)"; "Given that other people will probably do worse things at the rally, these behaviors are not that bad; and "As long as other people do worse things at the rally, these behaviors are acceptable." For euphemistic labeling they were also presented with the stem "To what extent do you think…" followed by the completions "These behaviors are just part of the fun at rallies"; "These behaviors are more of a joke than serious"; "These behaviors merely reflect healthy competition between people attending rallies"; "It is important to take these behaviors seriously (R)"; "These behaviors are no laughing matter (R)"; "People who take issue with these kinds of behaviors can't take a joke"; and "These behaviors make rallies exciting, like a sport." For denial of consequences they read: "No one would actually be harmed by these behaviors"; "These behaviors won't have lasting consequences"; "No one would be seriously affected by these behaviors"; "These behaviors can be harmful (R)"; "The consequences of behaviors like these can be severe (R)"; "The harms associated with these behaviors can negatively affect a person for a long time (R)"; and "These behaviors are not physically or psychologically harmful." For dehumanization they saw an adapted version of the denial of Human Uniqueness subscale (Bastian & Haslam, 2010; Bastian, Denson, & Haslam, 2013): "[pipe in opposing stance: Supporters/Opponents of multiculturalism] are barbaric"; " [pipe in opposing stance: Supporters/Opponents of multiculturalism] lack self-restraint, like animals"; "[pipe in opposing stance: Supporters/Opponents of multiculturalism] are unsophisticated"; "No one deserves to be treated like animals, not even [pipe in opposing stance: supporters/opponents of multiculturalism] (R)"; " [pipe in opposing stance: Supporters/Opponents of multiculturalism] are refined and cultured (R)"; " [pipe in opposing stance: Supporters/Opponents of multiculturalism] are rational and logical, like they are intelligent (R)"; and " [pipe in opposing

stance: Supporters/Opponents of multiculturalism] are less than human, like animals." For each item, they were presented with the response options *not at all, slightly, moderately, much,* and *very much.*

Worldview threat. Another aim of the Pilot Study was to pretest the items measuring worldview threat perceived at the rally, a potential moderator of my model of normative disengagement. Embedded in the block of questions examining factors driving harm endorsement, participants were asked the extent to which they perceived the upcoming rally as threatening to their worldview. Participants were presented with the stem: "To what extent do you think…" followed by seven items randomized within the 28 justification for harm items described above: "the rally will violate your core political values and beliefs"; "your political values and beliefs will be undermined at the upcoming rally"; "the rally is dangerous for society"; "the upcoming rally will threaten democracy in the United States"; "American values are at stake at the upcoming rally"; "The upcoming rally is a threat to American culture"; and "The upcoming rally will make American society more dangerous." Response options included *not at all*, *slightly*, *moderately*, *much*, and *very much*.

Attitudes toward multiculturalism. First, participants were provided a definition of multiculturalism: "Multiculturalism means that all different groups within society are treated equally and have the same rights, and their cultural perspectives are given equal value and status (e.g., Muslims/Christians, Blacks/Whites, gay and straight people, etc.)." Participants were asked to report the extent to which they perceived that multiculturalism is a key issue involved in Alt-Right/Antifa rallies: "To what extent do you perceive **multiculturalism** to be a key issue at stake at **Alt-Right/Antifa rallies**?" Response options included *not at all*, *slightly*, *moderately*, *much*, and *very much*. Second, participants were asked to report the extent to which they perceived that

support/opposition to multiculturalism is associated with liberals/conservatives. Specifically, participants were asked: "To what extent do you perceive **support for multiculturalism** as a liberal or conservative position?" and "To what extent do you perceive **opposition to multiculturalism** as a liberal or conservative position?" with response choices *very liberal*, *moderately liberal*, *slightly liberal*, *neither liberal nor conservative*, *slightly conservative*, *moderately conservative*, and *very conservative*.

Results

Harmfulness Pilot

The purpose of the harmfulness version of the Pilot was to confirm that the behaviors listed in Appendix A were perceived as harmful to some degree (i.e., greater than 1/*not at all harmful* in one-sample *t*-tests).

Preliminary analyses.

Bot detection. One hundred percent of participants ($N = 110$) passed the Captcha question, suggesting that none of them were bots.

Attention checks. I administered three self-reported close-ended questions assessing the degree to which participants were paying attention/not distracted, as well as one open-ended question designed to confirm that participants were reading questions carefully. Importantly, participants were reminded their answers to these questions would not impact their compensation in any way.

Participants self-reported on 5-point scales that I should *very much* trust their data ($M = 4.93$, $SD = 0.33$), that they were *not at all* distracted while completing the study ($M = 1.05$, $SD = 0.22$), and that they were *very much* paying attention during the study ($M = 4.96$, $SD = 0.20$). No participants fell below a mean level of *moderately* attentive, and I therefore did not filter

respondents on the basis of these self-reported questions. Corroborating these findings, 86.4% of

participants correctly identified their favorite movie when asked the question, "What is the topic

of this study? Sometimes participants do not carefully read the instructions. To correctly answer

this question, please select the option "other" and write down the name of your favorite movie."

In contrast, 3.6% of participants incorrectly selected "multiculturalism," and 10.0% skipped the

question.

Overall, these measures suggest participants were very attentive. Results described below

did not vary as a function of whether I excluded participants who failed the open-end question.

Therefore, I retained all participants for the analyses that follow.

Primary analyses.

The main goal of the harmfulness pilot study was to identify a corpus of behaviors that

people perceived as harmful to some degree. To identify behaviors that are harmful, I *a priori*

decided to discard any that did not significantly vary from "not at all" on the harmfulness scale

(i.e., a 1 on a 5-point scale) using one-sample *t*-tests.

As can be seen in Table 3, all Appendix A behaviors were perceived as significantly

harmful to some degree and were therefore retained for the endorsement version of the Pilot.

Table 3

Harmfulness Ratings (1-5 scale) and One-Sample t-Tests (test value = 1)

Behavior	M	SD	t	df	p	Lower 95% CI	Upper 95% CI
Fire	4.40	1.00	33.90	98	< .001	3.20	3.60
Punch	4.29	0.94	34.87	98	< .001	3.11	3.48
Throw large objects	4.21	1.05	30.37	98	< .001	3.00	3.42
Pepper spray	4.09	1.01	30.42	98	< .001	2.89	3.29
Kick	4.03	1.06	28.34	98	< .001	2.82	3.24
Slap	3.73	1.13	24.00	98	< .001	2.50	2.95
Vandalize	3.63	1.17	22.25	98	< .001	2.39	2.86
Push hard	3.63	1.10	23.70	98	< .001	2.41	2.85
Throw small objects	3.46	1.13	21.75	98	< .001	2.24	2.69
Smoke bomb	3.35	1.24	18.89	98	< .001	2.11	2.60
Spit	3.34	1.26	18.46	98	< .001	2.09	2.60
Push lightly	3.00	1.17	17.02	98	< .001	1.77	2.23
Barricades	2.42	1.33	10.69	98	< .001	1.16	1.69
Scream from close proximity	2.39	1.26	11.01	98	< .001	1.14	1.65
Human chain	2.23	1.19	10.35	98	< .001	1.00	1.47
Insults from close proximity	2.23	1.17	10.50	98	< .001	1.00	1.47
Offensive symbols	2.20	1.20	10.01	98	< .001	0.96	1.44
Swear from close proximity	2.11	1.20	9.19	98	< .001	0.87	1.35
Insults from a distance	2.02	1.13	8.95	98	< .001	0.79	1.25
Scream from a distance	1.99	1.11	8.86	98	< .001	0.77	1.21
Swear from a distance	1.93	1.15	8.01	98	< .001	0.70	1.16
Middle finger	1.87	1.11	7.77	98	< .001	0.65	1.09
Chanting in unison	1.79	1.08	7.25	98	< .001	0.57	1.00
Sit in	1.51	0.92	5.47	98	< .001	0.32	0.69
Picket	1.45	0.95	4.76	98	< .001	0.27	0.64
Hunger strike	1.42	0.96	4.40	98	< .001	0.23	0.62

Endorsement/Justifications Pilot

The purpose of the endorsement/justifications version of the Pilot was to confirm that the behaviors listed in Appendix A were endorsed to some degree (i.e., greater than 1/*not at all endorsed* in one-sample *t*-tests). Another aim of this version of the Pilot was to confirm that the justifications subscales were reliable, including the worldview threat subscale. My final goal in this Pilot was to conduct a factor analysis on the justification items to explore whether they factor in the way Bandura (1999) theorized.

Preliminary analyses.

Bot detection. All but one participant ($N = 223$) passed the Captcha question. The one participant who failed the Captcha was excluded from all analyses below.

Attention checks. I administered the same attention check measures as the harmfulness version of the Pilot.

Once again, participants self-reported on 5-point scales that I should *very much* trust their data ($M = 4.77$, $SD = 0.58$), that they were *not at all* distracted while completing the study ($M = 1.23$, $SD = 0.77$), and that they were *very much* paying attention during the study ($M = 4.88$, $SD = 0.37$). No participants fell below a mean level of *moderately* attentive, and I therefore did not filter respondents on the basis of these self-report questions. Corroborating these findings, 77.6% of participants correctly identified their favorite movie for the open-ended attention check. In contrast, 10.3% of participants incorrectly selected "multiculturalism," and 12.1% skipped the question.

Overall, these measures suggest participants were very attentive. Results described below were similar regardless of whether I excluded participants who failed the open-end question. See Tables 4 and 5 for a comparison of how results slightly shift based on exclusion decisions.

Whenever results varied as a function of exclusion decisions, I erred on the side of conservatism and chose stimuli that had stable results regardless of exclusion decisions. [3]

Primary analyses.

Endorsement ratings.

One goal of the Pilot Study was to identify a corpus of behaviors that people perceived as tempting to endorse to some degree. To identify behaviors that are endorsed to some degree, I *a priori* decided to discard any that did not significantly vary from "not at all" on the endorsement scale (i.e., a 1 on a 5-point scale) using one-sample *t*-tests. As can be seen in Tables 4 and 5, all Appendix A behaviors were perceived as significantly endorsed to some degree and were therefore retained.

A second goal of the Pilot was to ensure that any retained behaviors were uncorrelated with political orientation. As can be seen in Tables 4 and 5, 11 behaviors were consistently uncorrelated with political orientation regardless of my exclusion decisions, and they were therefore retained for use in the Main Studies. The 15 behaviors that were correlated with political orientation were discarded. See Table 6 for a scorecard of how all behaviors fared in light of decision rules surrounding harmfulness, endorsement, and correlation with political orientation.

Factor analysis of justification items.

The following results did not vary as a function of whether participants passed the open-ended attention check, so I retained all of them for these analyses.

---

[3] Analyses filtering out those who failed the attention check are included in Table 5. Comparing these results to Table 4, the levels of significance changed for four correlations on this basis of including/excluding these participants. I proceeded and retained only those behaviors that were uncorrelated with political orientation regardless of my decision to retain/exclude participants.

I conducted a confirmatory factor analysis to test a four-factor model of justifications for harm (i.e., advantageous comparison, euphemistic labeling, denial of consequences, and dehumanization). All factors were measured with seven items each as discussed in the Method section.

I fit the model with the lavaan package version 0.6-3 (Rosseel, 2012) in R using maximum likelihood estimation. The model fit was not acceptable, with a Tucker-Lewis Index (TLI) of 0.60 (>0.90 is acceptable) and a RMSEA of 0.15 [90% CI: 0.143, 0.158] (<0.08 is adequate). I attempted to improve model fit by removing reverse-coded items from the four-factor model. The model fit was still not acceptable (TLI = 0.87, RMSA = 0.11).

Given that confirmatory factor analysis models did not adequately fit the data, I proceeded by conducting a data-driven principal components analysis with varimax rotation. Entering all items into the model yielded an uninterpretable seven-factor solution (see Table 7). As can be seen in Table 7, the reverse-coded items from each subscale in particular did not load cleanly with regularly worded items. Based on this observation, I ran another principal components analysis with varimax rotation excluding all reverse-coded items. After I removed reverse-coded items, a four-factor solution emerged that largely resembled Bandura's (1999) mechanisms of reprehensible behavior (See Table 8). In rare cases, some items loaded onto the wrong factor or cross-loaded onto multiple factors. These items were discarded for the Main Studies (highlighted in gray in Table 8). I wrote two additional advantageous comparison items ("Other people will be more obnoxious than this" and "Other people will be more harmful than this") to ensure that I had at least four items for each subscale before turning to the Main Studies.

Reliability of justification subscales.

The final goal of this Pilot was to confirm that the retained items from the justification subscales were reliable. Results do not vary as a function of whether participants passed or failed the open-ended attention check, so all participants were retained for these analyses. After removing reversed-coded and poor loading items, as suggested by the factor analysis, the advantageous labeling ($r = 0.80$), euphemistic labeling ($\alpha = 0.88$), denial of consequences ($\alpha = 0.91$), dehumanization ($\alpha = 0.89$), and threat ($\alpha = 0.89$) subscales were reliable.

Table 4

Endorsement Ratings, One-Sample t-Tests (test value = 1), & Correlations with Political Orientation (ALL PARTICIPANTS)

Behavior	*M*	*SD*	*t*	df	*p*	Lower 95% CI	Upper 95% CI	*r* (with PO)
Picket	3.20	1.53	20.48	203	< .001	1.98	2.41	-0.07
Sit in	2.75	1.53	16.32	203	< .001	1.54	1.96	-0.10
Chanting in unison	2.46	1.45	14.36	203	< .001	1.26	1.66	-0.27***
Human chain	2.01	1.35	10.68	203	< .001	0.82	1.20	-0.21**
Hunger strike	2.00	1.35	10.55	203	< .001	0.81	1.19	-0.15*
Scream from a distance	1.97	1.28	10.77	203	< .001	0.79	1.14	-0.20**
Swear from a distance	1.91	1.26	10.30	203	< .001	0.73	1.08	-0.26***
Middle finger	1.91	1.29	10.09	203	< .001	0.73	1.09	-0.21**
Insults from a distance	1.89	1.30	9.82	203	< .001	0.71	1.07	-0.22**
Offensive symbols	1.87	1.28	9.74	203	< .001	0.70	1.05	-0.21**
Insults from close proximity	1.85	1.28	9.47	203	< .001	0.67	1.02	-0.15*
Barricades	1.84	1.30	9.23	203	< .001	0.66	1.02	-0.24**
Swear from close proximity	1.83	1.23	9.62	203	< .001	0.66	1.00	-0.22**
Scream from close proximity	1.79	1.27	8.84	203	< .001	0.61	0.97	-0.15*
Push hard	1.50	1.14	6.31	203	< .001	0.35	0.66	-0.09
Throw small objects	1.49	1.18	5.89	203	< .001	0.32	0.65	-0.10
Kick	1.46	1.10	6.00	202	< .001	0.31	0.62	-0.11
Spit	1.44	1.03	6.11	203	< .001	0.30	0.58	-0.08
Vandalize	1.44	1.04	5.98	203	< .001	0.29	0.58	-0.16*
Pepper spray	1.43	1.03	6.00	203	< .001	0.29	0.57	-0.05
Slap	1.43	1.03	5.97	203	< .001	0.29	0.57	-0.14*
Smoke bomb	1.43	1.01	6.02	203	< .001	0.29	0.57	-0.11
Push lightly	1.42	0.97	6.23	203	< .001	0.29	0.56	-0.09
Throw large objects	1.41	1.03	5.71	203	< .001	0.27	0.55	-0.05
Punch	1.39	0.95	5.91	203	< .001	0.26	0.52	-0.09
Fire	1.38	1.03	5.23	203	< .001	0.24	0.52	-0.11

Note. * *p* < .05; ** *p* < .01; *** *p* < .001. PO: Lower values = More liberal.

Table 5

Endorsement Ratings, One-Sample t-Tests (test value = 1), & Correlations with Political Orientation (EXCLUDING INATTENTIVES)

Behavior	M	SD	t	df	p	Lower 95% CI	Upper 95% CI	r (with PO)
Picket	3.23	1.54	19.08	172	< .001	2.00	2.46	-0.12
Sit in	2.86	1.53	15.96	172	< .001	1.63	2.09	-0.10
Chanting in unison	2.49	1.48	13.19	172	< .001	1.26	1.71	-0.30***
Human chain	2.01	1.38	9.67	172	< .001	0.81	1.22	-0.22**
Hunger strike	2.06	1.39	10.02	172	< .001	0.85	1.27	**-0.14**
Scream from a distance	1.99	1.28	10.15	172	< .001	0.80	1.18	-0.19*
Swear from a distance	1.91	1.27	9.45	172	< .001	0.72	1.10	-0.27***
Middle finger	1.93	1.30	9.41	172	< .001	0.74	1.13	-0.23**
Insults from a distance	1.90	1.30	9.11	172	< .001	0.71	1.10	-0.24**
Offensive symbols	1.87	1.28	8.88	172	< .001	0.67	1.06	-0.23**
Insults from close proximity	1.84	1.27	8.68	172	< .001	0.65	1.03	-0.16*
Barricades	1.87	1.32	8.64	172	< .001	0.67	1.07	-0.28***
Swear from close proximity	1.84	1.26	8.78	172	< .001	0.65	1.03	-0.23**
Scream from close proximity	1.76	1.26	7.99	172	< .001	0.57	0.95	-0.17*
Push hard	1.50	1.13	5.77	172	< .001	0.33	0.67	-0.07
Throw small objects	1.47	1.16	5.38	172	< .001	0.30	0.65	-0.09
Kick	1.45	1.10	5.41	171	< .001	0.29	0.62	-0.11
Spit	1.45	1.03	5.68	172	< .001	0.29	0.60	-0.07
Vandalize	1.43	1.02	5.53	172	< .001	0.27	0.58	**-0.14**
Pepper spray	1.41	0.99	5.46	172	< .001	0.26	0.56	-0.06
Slap	1.43	1.04	5.41	172	< .001	0.27	0.58	**-0.13**
Smoke bomb	1.38	0.95	5.26	172	< .001	0.24	0.52	**-0.16***
Push lightly	1.42	0.96	5.68	172	< .001	0.27	0.56	-0.12
Throw large objects	1.40	1.02	5.21	172	< .001	0.25	0.56	-0.04
Punch	1.38	0.94	5.36	172	< .001	0.24	0.52	-0.07
Fire	1.38	1.03	4.87	172	< .001	0.23	0.54	-0.15

Note. * *p* < .05; ** *p* < .01; *** *p* < .001. Bold indicates a change in significance from Table 4. PO: Lower values = More liberal.

Table 6

Pilot Study Scorecard

Behavior	Harmful	Endorsed	Uncorrelated with PO
Picket	✓	✓	✓
Sit in	✓	✓	✓
Push hard	✓	✓	✓
Throw small objects	✓	✓	✓
Kick	✓	✓	✓
Spit	✓	✓	✓
Pepper Spray	✓	✓	✓
Push lightly	✓	✓	✓
Throw large objects	✓	✓	✓
Punch	✓	✓	✓
Fire	✓	✓	✓
Chanting in unison	✓	✓	
Human chain	✓	✓	
Hunger strike	✓	✓	
Scream from a distance	✓	✓	
Scream from close proximity	✓	✓	
Swear from a distance	✓	✓	
Swear from close proximity	✓	✓	
Insults from a distance	✓	✓	
Insults from close proximity	✓	✓	
Middle finger	✓	✓	
Offensive symbols	✓	✓	
Barricades	✓	✓	
Vandalize	✓	✓	
Slap	✓	✓	
Smoke bomb	✓	✓	

Table 7

Principal Components Analysis of Justifications (ALL ITEMS)

	Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Comp1	Other protesters would normally behave worse than this	0.23	0.07	-0.05	0.05	**0.89**	-0.07	-0.10
Comp2	Other people will do worse things than this at the upcoming rally	0.10	0.00	-0.12	0.15	**0.89**	-0.11	-0.10
Comp3	These behaviors are better than how protesters normally act at rallies	**0.51**	0.26	-0.01	0.02	**0.45**	-0.02	0.27
Comp4R	No one else I know would behave in these ways	-0.05	0.08	0.24	-0.12	-0.09	0.17	**0.75**
Comp5R	These behaviors are worse than how people normally act at rallies	-0.18	-0.01	0.29	-0.21	-0.03	0.21	**0.63**
Comp6	Given that other people will probably do worse things at the rally, these behaviors are not that bad	**0.42**	**0.40**	0.03	0.18	**0.35**	-0.25	0.27
Comp7	As long as other people do worse things at the rally, these behaviors are acceptable	**0.67**	0.31	0.10	0.18	0.23	-0.16	0.10
DenCon1	No one would actually be harmed by these behaviors	0.23	**0.83**	0.22	0.03	0.09	-0.07	0.03
DenCon2	These behaviors won't have lasting consequences	0.33	**0.75**	0.18	0.08	-0.02	-0.11	-0.01
DenCon3	No one would be seriously affected by these behaviors	0.32	**0.83**	0.18	0.05	0.07	-0.14	-0.01
DenCon4R	These behaviors can be harmful	0.02	0.29	**0.78**	0.02	0.05	-0.08	0.28
DenCon5R	The consequences of behaviors like these can be severe	-0.04	0.26	**0.82**	-0.02	0.01	-0.02	0.28
DenCon6R	The harms associated with these behaviors can negatively affect a person for a long time	0.04	0.22	**0.79**	0.02	0.03	0.01	0.30

DenCon7	These behaviors are not physically or psychologically harmful	0.25	**0.84**	0.04	0.05	-0.01	-0.06	-0.04
Dehum1	[Targets] are barbaric	0.01	0.02	-0.07	**0.90**	0.09	0.11	-0.03
Dehum2	[Targets] lack self-restraint, like animals	0.11	0.03	-0.06	**0.92**	0.03	0.06	-0.06
Dehum3	[Targets] are unsophisticated	0.17	0.08	-0.06	**0.85**	0.04	0.17	-0.04
Dehum4R	No one deserves to be treated like animals, not even [targets]	0.28	-0.16	0.31	0.15	-0.04	-0.09	**0.64**
Dehum5R	[Targets] are refined and cultured	-0.21	-0.15	0.02	0.11	-0.11	**0.87**	0.16
Dehum6R	[Targets] are rational and logical, like they are intelligent	-0.30	-0.17	0.10	0.13	-0.11	**0.81**	0.10
Dehum7	[Targets] are less than human, like animals	0.31	0.13	-0.01	**0.64**	0.12	-0.22	-0.02
Euph1	These behaviors are just part of the fun at rallies	**0.77**	0.33	0.13	0.11	0.16	-0.06	-0.09
Euph2	These behaviors are more of a joke than serious	**0.79**	0.19	0.21	0.06	0.06	-0.14	-0.17
Euph3	These behaviors merely reflect healthy competition between people attending rallies	**0.56**	**0.51**	0.13	0.14	0.23	-0.02	0.06
Euph4R	It is important to take these behaviors seriously	0.34	-0.08	**0.75**	-0.14	-0.18	0.14	-0.01
Euph5R	These behaviors are no laughing matter	0.28	0.05	**0.81**	-0.14	-0.13	0.09	0.03
Euph6	People who take issue with these kinds of behaviors can't take a joke	**0.73**	0.19	0.03	0.20	-0.02	-0.21	0.07
Euph7	These behaviors make rallies exciting, like a sport	**0.75**	0.32	0.12	0.14	0.15	-0.19	-0.02
	Eigenvalue	8.45	4.54	2.79	1.70	1.51	1.03	1.01
	% variance explained	30.17%	16.21%	9.96%	6.08%	5.39%	3.66%	3.62%

Table 8

Principal Components Analysis of Justifications (REMOVING REVERSE-CODED ITEMS)

	Item	Factor 1	Factor 2	Factor 3	Factor 4
Comp1	Other protesters would normally behave worse than this	0.19	0.07	0.06	**0.91**
Comp2	Other people will do worse things than this at the upcoming rally	0.07	-0.01	0.14	**0.91**
Comp3	These behaviors are better than how protesters normally act at rallies	**0.48**	0.26	0.02	**0.47**
Comp6	Given that other people will probably do worse things at the rally, these behaviors are not that bad	**0.43**	**0.43**	0.13	**0.39**
Comp7	As long as other people do worse things at the rally, these behaviors are acceptable	**0.71**	0.31	0.16	0.24
DenCon1	No one would actually be harmed by these behaviors	0.25	**0.86**	0.02	0.09
DenCon2	These behaviors won't have lasting consequences	0.31	**0.81**	0.06	0.00
DenCon3	No one would be seriously affected by these behaviors	0.33	**0.87**	0.03	0.08
DenCon7	These behaviors are not physically or psychologically harmful	0.24	**0.83**	0.06	0.02
Dehum1	[Targets] are barbaric	0.00	0.00	**0.92**	0.08
Dehum2	[Targets] lack self-restraint, like animals	0.09	0.04	**0.93**	0.05
Dehum3	[Targets] are unsophisticated	0.14	0.06	**0.88**	0.03
Dehum7	[Targets] are less than human, like animals	0.36	0.13	**0.62**	0.16
Euph1	These behaviors are just part of the fun at rallies	**0.80**	0.31	0.11	0.15

Euph2	These behaviors are more of a joke than serious	**0.83**	0.20	0.05	0.06
Euph3	These behaviors merely reflect healthy competition between people attending rallies	**0.58**	**0.50**	0.14	0.22
Euph6	People who take issue with these kinds of behaviors can't take a joke	**0.75**	0.20	0.18	0.03
Euph7	These behaviors make rallies exciting, like a sport	**0.80**	0.32	0.11	0.16
	Eigenvalue	7.58	2.73	1.78	1.19
	% variance explained	42.11%	15.15%	9.91%	6.58%

Note. Items highlighted in gray indicate items that loaded on the wrong factor or cross-loaded on multiple factors.

Combined Pilot

Lastly, I pooled together both versions of the Pilot to assess perceptions of a video from a past Alt-Right rally (i.e., the 2017 Unite the Right rally) and perceptions of multiculturalism. Specifically, I was expecting to find that the behavior depicted in the video was perceived as at least *slightly* violent and obnoxious, that the issue of multiculturalism was perceived as at least *slightly* relevant to Alt-Right/Antifa rallies, and that support for (opposition to) multiculturalism was perceived as a liberal (conservative) position. All of these hypotheses were tested using one sample *t*-tests. None of the results below varied as a function of whether participants passed or failed the open-ended attention check, so I retained them all for these analyses.

Perceptions of the video.

The video was seen as significantly more violent than the *slightly* point of the five-point scale ($M = 3.85$, $SD = 0.95$), $t(308) = 34.16$, $p < .001$, 95% CI: [1.75, 1.96]. The behavior in the video was also seen as significantly more obnoxious than the *slightly* point of the five-point scale ($M = 4.20$, $SD = 0.99$), $t(308) = 39.24$, $p < .001$, 95% CI: [2.09, 2.31]. Taken together, these results suggest that the video is suitable for use in the Main Studies given that it is perceived as violent and obnoxious. Therefore, when I refer to the video as displaying violent and obnoxious behavior, Main Study participants will agree with that description.

Perceptions of multiculturalism.

As expected, multiculturalism was perceived as at least *slightly* relevant to Alt-Right/Antifa rallies on a five-point scale of relevance ($M = 3.65$, $SD = 1.21$), $t(294) = 23.47$, $p < .001$, 95% CI: [1.52, 1.79].

Moreover, as anticipated, support for multiculturalism ($M = -1.49$, $SD = 1.48$) was seen as significantly more liberal than the politically neutral point (i.e., 0 on a scale ranging from -3

(*very liberal*) to +3 (*very conservative*)), $t(294) = -17.33$, $p < .001$, 95% CI: [-1.66, -1.32]. Similarly, opposition to multiculturalism ($M = 1.57$, $SD = 1.59$) was perceived as significantly more conservative than the politically neutral point, $t(294) = 16.93$, $p < .001$, 95% CI: [1.38, 1.75].

Taken together, these results reveal that multiculturalism was an appropriate attitude object to assess in the context of Alt-Right/Antifa rallies, given its perceived relevance to the rallies. Moreover, support for (opposition to) multiculturalism was spontaneously perceived as a liberal (conservative) position, which ensured that supporters/opposers of the issue were perceived as liberals/conservatives in the Main Studies.

STUDIES 1 & 2

Studies 1 and 2 were designed to explore the various factors that drive willingness to harm obnoxious protesters at a political rally, including the extent to which the effect is symmetrical (or asymmetrical) across the political divide and moderated by perceptions of worldview threat. I collected all data for Studies 1 and 2 at one time point and within one sample (manipulating the targets of evaluation between-subjects: obnoxious opponents of multiculturalism vs. obnoxious supporters of multiculturalism). For simplicity, however, I broke my analyses into two main parts: one part was called Study 1 and the other was called Study 2. In Study 1, I explored factors that drove harm endorsement toward opponents of multiculturalism for participants of all stances toward the issue. In Study 2, I explored these patterns toward supporters of multiculturalism, using participants of all stances. Finally, I did an aggregated analysis to examine whether attitudinal congruence with the target, respondents' political ideology (whether they had a conservative or liberal position on the issue of multiculturalism), and participants' moral conviction for multiculturalism interacted to predict willingness to harm opponents at a political rally.

Method

Participants

Eight hundred seventy-three workers from Amazon's Mechanical Turk (U.S. only; HIT approval rate > 90%) were recruited to participate in an online study. This sample size was estimated using G*Power assuming small effects and 80% power.[4] In return for their participation, participants were paid $0.10/minute ($1.40 total). Forty-two percent of participants identified as female, and on average, were 38.50 years of age ($SD = 12.16$). Most participants identified as White (72.2%), followed by Asian/Pacific Islander (6.8%), Latino (6.6%), African American (5.4%), Native American (2.1%) and Other (0.9%). 43.6% of participants attained a Bachelor's degree or higher; 33.7% of participants completed some college or a 2-year college degree; 11.8% of participants never attended college. Participants were on average politically neutral ($M = 0.18$, $SD = 2.75$) on a scale of (-4) *very much liberal* to (+4) *very much conservative*. See Table 9 for a breakdown of participants' political orientation.

Table 9
Main Study: Participants' Political Orientation

Political Orientation	Frequency	%
Very much liberal	111	13.7
Much liberal	86	10.6
Moderately liberal	77	9.5
Slightly/lean liberal	62	7.7
Neutral/neither	60	7.4
Slightly/lean conservative	89	11.0
Moderately conservative	99	12.2
Much conservative	122	15.1
Very much conservative	103	12.7

---

[4] Regression estimates in G*Power assume within-subjects designs, but I will manipulate the target of potential harm endorsement between-subjects. To get around this issue, I used the ANCOVA calculator in G*Power using target of potential harm endorsement as a between-subjects factor (2 groups: conservative targets vs. liberal targets). I then included all other measured variables as covariates to yield a sample size estimate of approximately 800 assuming small effects and 80% power.

Procedure

Participants entered the survey and completed the consent process, as well as a Captcha question (see Appendices B & C) to confirm that they were human participants. Participants were then asked to report their attitudes toward multiculturalism, including their support/opposition to it, the degree to which the attitude was important and certain, and the degree to which they viewed the issue through the lens of moral conviction (see Measures section below). Like the Pilot Study, only participants who reported that they supported or opposed the issue to some degree were invited to complete the remainder of the study (otherwise I could not determine whether they were attitudinally congruent/incongruent with targets who supported/opposed multiculturalism). Participants were also asked to report their political orientation. Participants were then presented with the description of a contentious Alt-Right political rally that is poised to take place in 2019. They watched the same video as the Pilot Study about the conflict that characterized the 2017 Unite the Right Rally.

On the next page, they were told: "*Imagine you attend the upcoming Alt-Right/Antifa rally. What kind of actions would you be comfortable or uncomfortable with your side taking against [manipulated between-subjects: **opponents vs. supporters of multiculturalism**] who display obnoxious behavior like that depicted in the video? What actions would you **support or oppose** using against [**opponents vs. supporters of multiculturalism**] who attend the rally and display obnoxious behavior like that depicted in the video?*"
Participants completed a measure that evaluated their personal endorsement of harmful behaviors enacted against the obnoxious target protesters. Specifically, they read, "To what extent would you **support or oppose** using the following actions against obnoxious [**opponents vs. supporters of multiculturalism**] at the upcoming rally?" followed by the top eight most

endorsed behaviors retained from the Pilot Study.[5] Response options included *strongly support,*

moderately support, *slightly support*, *neither support nor oppose, slightly oppose*, *moderately*

oppose, and *strongly oppose*.

As a secondary measure of endorsement of harmful actions against opposing protesters, I

asked participants to evaluate each action again, but this time asked them the extent to which

they would encourage or discourage *others* who were doing the actions. Specifically, they read,

"To what extent would you **encourage or discourage** others who use the following actions

against obnoxious [**opponents vs. supporters of multiculturalism**] at the upcoming rally?"

Response options included *strongly encourage, moderately encourage, slightly encourage,*

neither encourage nor discourage, *slightly discourage*, *moderately discourage*, *strongly*

discourage.

As a third measure of harm endorsement, I asked participants to evaluate each action

again, but this time asked them the extent to which they found each action justifiable or

unjustifiable. Specifically, they read, "To what extent do you think the following actions are

justifiable or unjustifiable to use against obnoxious [**opponents vs. supporters of**

multiculturalism] at the upcoming rally?" Response options included *very justifiable,*

moderately justifiable, *slightly justifiable*, *neither justifiable nor unjustifiable*, *slightly*

unjustifiable, moderately unjustifiable, and *very unjustifiable.*

Next, to examine the extent to which of Bandura's (1999) routes to reprehensible

behavior drove harm endorsement (or weakened opposition to harm), I next asked people to

consider various factors that influenced their judgments. Specifically, they read the prompt,

---

[5] This choice was driven by a desire to limit the time it would take to complete the Main Study
sessions and to limit participant fatigue.

"Next, we are interested in learning more about how you decided which behaviors to

support or oppose in the previous task. Thinking about your responses overall, how much

do you agree that each factor below influenced your decision to support/oppose these

behaviors?"

They were then presented with a series of statements inspired by Bandura's (1999)

mechanisms. Embedded in these items was a measure of the extent to which they perceived the

rally to be threatening to their worldview. See Measures section below for more detail.

Next, to assess the impact of their endorsements on their moral self-image, they

completed a moral self-image scale (Jordan, Leliveld, & Tenbrunsel, 2015). To ensure that

participants could not guess the purpose of this measure, target moral self-image items were

embedded among other non-moral traits (e.g., warmth and competence), and participants were

told the purpose of this measure was to learn more about their personalities. See more details in

the Measures section below.

To examine the extent to which participants perceived harming others as a normative

standard that could be flexibly broken when a strong moral conviction was at stake, I

administered an adapted version of the idealism subscale of Forsyth's (1980) Ethics Position

Questionnaire. If my theory is correct, I expected to find that higher moral conviction for

multiculturalism would be significantly associated with weaker ethical idealism scores. See more

details in the Measures section below.

Participants then completed a manipulation check by completing the following question:

"In this study you were asked to report the extent to which you would support/oppose enacting

various behaviors toward protesters at upcoming Alt-Right/Antifa rally. What protesters did you

evaluate during this task?" Response options included: *supporters of multiculturalism*, *opponents*

of multiculturalism, another group of protesters not listed here. Participants passed the

manipulation check if they successfully identified the target type that they were randomly

assigned to evaluate.

During the last block of questions, participants reported on the extent to which we should

trust their data (to ensure that they were paying attention). Participants were asked: "To what

extent should we trust your data?"; "How distracted were you while completing this study? (R)";

and "To what extent were you paying attention in this study?" Response options for all these

questions included *not at all*, *slightly*, *moderately*, *much*, and *very much*. As a final question to

make sure participants were paying attention, they were asked: "What is the topic of this study?

Sometimes participants do not carefully read the instructions. To correctly answer this question,

please select the option "other" and write down the name of your favorite movie." Response

options will include *sexism*, *multiculturalism*, *I don't remember*, and *other* with a text-entry field

next to it. Participants passed this attention check if they wrote a movie title in the "other" field.

At the conclusion of the study, participants were thanked for their time and compensated.

Measures

Support for/opposition to multiculturalism. Support/opposition to multiculturalism

was assessed using the same items/answer choices as the Pilot Study.

On average, participants were moderately supportive of multiculturalism (M = -2.02, SD

= 2.30). See Table 10 for a breakdown of participants' attitudes toward multiculturalism.

Table 10
Main Study: Participants' Attitudes toward Multiculturalism

Attitude Position Toward Multiculturalism	Frequency	%
Very much support	311	36.4
Much support	170	19.9
Moderately support	121	14.2
Slightly/lean support	58	6.8
Neutral/neither	40	4.7
Slightly/lean oppose	70	8.2
Moderately oppose	21	2.5
Much oppose	31	3.6
Very much oppose	32	3.7

Attitude congruence with the target. Based on participants' measured attitude positions about multiculturalism, they were categorized as supporters of multiculturalism or opponents of multiculturalism. In Study 1 in which the targets were obnoxious opponents of multiculturalism, supporters of multiculturalism were categorized as attitudinally incongruent with the target, whereas opponents were attitudinally congruent with them. In contrast, in Study 2 in which the targets were obnoxious supporters of multiculturalism, supporters of the issue were attitudinally congruent with the targets, but opponents were incongruent with them.[6]

Attitude strength for multiculturalism. Participants were reminded of the definition of multiculturalism. They were then presented with the stem, "To what extent is your position on **multiculturalism in the U.S.**…" followed by four completions: "something that you care a lot about?"; "personally important to you?"; "something you are certain about?"; and "something

---

[6] Although the decision to dichotomize a continuous variable is generally discouraged, I decided that it was appropriate for analyses below because: (1) it simplified complex patterns of results; (2) results did not markedly vary as a function of this decision; (3) I hypothesized about the effects of being attitudinally congruent vs. incongruent with the speaker in a dichotomous fashion rather than as a continuous variable.

you are sure you are correct about?" Participants responded on five-point scales, with the point labels of *not at all*, *slightly*, *moderately*, *much*, and *very much*. Participants' responses to these items were averaged to compute an overall attitude strength score. The attitude importance items ("something that you care a lot about?"; "personally important to you?") were reliable ($r = .82$), as were the attitude certainty items ("something you are certain about?"; "something you are sure you are correct about?"), $r = 0.77$.

  Moral conviction for multiculturalism. After presenting participants with a definition of multiculturalism, they were presented with the stem, "To what extent is your position on **multiculturalism in the U.S.**…" followed by four completions: "…a reflection of your core moral beliefs and convictions?"; "…connected to your beliefs about fundamental right and wrong?"; "…based on moral principle?"; "…a moral stance?" They were presented with the response options *not at all*, *slightly*, *moderately*, *much*, and *very much*. Participants' responses to these items were averaged to compute an overall moral conviction score. The moral conviction items were reliable, $\alpha = 0.92$.

  Political orientation. Political orientation was assessed with the same items and answer choices as the measure in the Pilot Study.

  Factors driving harm endorsement. To examine the extent to which each of Bandura's (1999) theorized mechanisms drove harm endorsement, I utilized the retained items and answer choices from the Pilot Study.

  Corroborating the Pilot studies, the advantageous comparison ($\alpha = 0.93$), euphemistic labeling ($\alpha = 0.85$), denial of consequences ($\alpha = 0.88$), and dehumanization ($\alpha = 0.92$) subscales were reliable. Moreover, a principal components analysis revealed that these items factored separately according to Bandura's justifications (see Table 11).

Table 11

Principal Components Analysis of Justifications: Main Study

	Item	Factor 1	Factor 2	Factor 3	Factor 4
Comp1	Other protesters would normally behave worse than this	**0.85**	0.14	0.08	0.15
Comp2	Other people will do worse things than this at the upcoming rally	**0.90**	0.13	0.03	0.13
Comp3	Other people will be more obnoxious than this	**0.90**	0.15	0.00	0.06
Comp4	Other people will be more harmful than this	**0.90**	0.15	0.04	0.07
Dehum1	[Targets] are barbaric	0.16	**0.90**	0.02	0.09
Dehum2	[Targets] lack self-restraint, like animals	0.15	**0.91**	0.07	0.12
Dehum3	[Targets] are unsophisticated	0.16	**0.89**	0.02	0.07
Dehum4	[Targets] are less than human, like animals	0.12	**0.80**	0.11	0.25
DenCon1	No one would actually be harmed by these behaviors	0.00	0.03	**0.82**	0.25
DenCon2	These behaviors won't have lasting consequences	0.08	0.04	**0.80**	0.19
DenCon3	No one would be seriously affected by these behaviors	0.04	0.06	**0.86**	0.27
DenCon4	These behaviors are not physically or psychologically harmful	0.03	0.07	**0.82**	0.22
Euph1	These behaviors are just part of the fun at rallies	0.09	0.15	0.29	**0.75**
Euph2	These behaviors are more of a joke than serious	0.09	0.16	0.24	**0.79**
Euph3	People who take issue with these kinds of behaviors can't take a joke	0.15	0.12	0.23	**0.74**

Euph4	These behaviors make rallies exciting, like a sport	0.09	0.09	0.22	**0.81**
	Eigenvalue	3.30	3.23	2.98	2.74
	% variance explained	20.61%	20.15%	18.63%	17.14%

Worldview threat. Embedded in the block of questions examining factors driving harm endorsement, participants were asked the extent to which they perceived the upcoming rally as threatening to their worldview. I utilized the same items and answer choices as the measure pretested in the Pilot Study. The threat ($\alpha = 0.90$) subscale was reliable.

Moral self-image. Participants' moral self-images were assessed using Jordan et al.'s (2015) Moral Self-Image Scale: a measure that captures people's dynamic fluctuations in the extent to which they perceive themselves as moral at a specific point in time. Participants read the instructions, "Please respond to the following statements as they apply to you" following by nine randomized statements: "Compared to the caring person I want to be, I am…"; "Compared to the compassionate person I want to be, I am…"; "Compared to the fair person I want to be, I am…"; "Compared to the generous person I want to be, I am…"; "Compared to the moral person I want to be, I am…"; "Compared to the ethical person I want to be, I am…"; "Compared to the hard-working person I want to be, I am…"; "Compared to the honest person I want to be, I am…"; "Compared to the loyal person I want to be, I am…"; "Compared to the respectful person I want to be, I am…". Response options ranged from -3 (*Much less [moral trait] than the person I want to be*) to +3 (*Much more [moral trait] than the person I want to be*), with the middle point of the scale 0 being *Exactly as [moral trait] as the person I want to be*. Participants' responses to these items were averaged to compute an overall moral self-image score. To mask the purpose of this scale, I also randomly interspersed non-moral traits traditionally related to warmth and competence dimensions of person perception (e.g., Leach, Ellemers, & Barreto, 2007). Specifically, these distractor items included: "Compared to the likeable person I want to be, I am…"; "Compared to the warm person I want to be, I am…"; "Compared to the friendly person I want to be, I am…"; "Compared to the competent person I want to be, I am…"; "Compared to

the intelligent person I want to be, I am…"; "Compared to the skilled person I want to be, I

am…". Response options again ranged from -3 (*Much less [trait] than the person I want to be*)

to +3 (*Much more [trait] than the person I want to be*), with the middle point of the scale 0 being

Exactly as [trait] as the person I want to be.

     Interestingly, a principal components analysis with varimax rotation revealed a single

factor solution that accounted for 56.15% of the variance in responding. Regardless, for

theoretical reasons, I proceeded below retaining only the items that were defined *a priori* as

morally relevant (Jordan et al., 2015). The morally relevant items comprising the moral self-

image scale were reliable ($\alpha = 0.91$).

     Ethical idealism. To examine the degree that participants perceived harmful political

behavior as an optional normative standard that could be broken in the service of a strong moral

conviction, I administered an adapted version of the idealism subscale of Forsyth's (1980) Ethics

Position Questionnaire.[7] Participants were presented with the directions: "You will find a series

of general statements listed below. Each represents a commonly held opinion and there are no

right or wrong answers. You will probably disagree with some items and agree with others. We

are interested in the extent to which you agree or disagree with such matters of opinion. Please

each statement carefully. Then indicate the extent to which you agree or disagree." They then

responded to the following questions: "A person should make certain that their political actions

associated with their beliefs about multiculturalism never intentionally harm another even to a

small degree"; "Harming another person with an opposing viewpoint on multiculturalism is

wrong, irrespective of how small the harms might be"; "Harming others to serve one's beliefs

---

[7] This scale was administered based on feedback from the proposal meeting to show that strong
moral conviction about multiculturalism negatively correlates with ethical idealism, that is, the
belief that it is never acceptable to harm others for a noble cause.

about multiculturalism is always wrong, irrespective of the benefits to be gained"; "One should

never psychologically or physically harm another person to advance his or her position on

multiculturalism"; "One should not perform an action which might in any way threaten the

dignity and welfare of another individual, even if that individual holds an opposing position on

the issue of multiculturalism"; "If a political action designed to serve beliefs about

multiculturalism could harm an innocent other, then it should not be done"; "Deciding whether

or not to perform an act by balancing the positive consequences of the act against the negative

consequences of the act is immoral, even if the act serves beliefs about multiculturalism"; "The

dignity and welfare of people should be the most important concern in any society, even for

people with opposing positions on the issue of multiculturalism"; and "It is never necessary to

sacrifice the welfare of others, even people with opposing positions on the issue of

multiculturalism". Response options included *strongly disagree*, *moderately disagree*, *slightly*

disagree, *neither agree nor disagree*, *slightly agree*, *moderately agree*, *strongly agree*. The

ethical idealism scale items were reliable ($\alpha = 0.89$).

Principal Components Analysis of Dependent Measures

I had three broad classes of dependent measures: support/opposition for eight harmful

protest behaviors toward randomly assigned targets, encouragement/ discouragement of others

who enact the eight behaviors toward randomly assigned targets, and justifiability/

unjustifiability of the eight behaviors toward randomly assigned targets. That is, each of the eight

harmful behaviors was evaluated three times by each participant, each time using a slightly

different wording. For simplicity, for each harmful behavior, I averaged across the three items to

yield one overall evaluation measure.

Next, I explored whether evaluations toward the eight harmful behaviors factored together. A principal components analysis with varimax rotation revealed that the eight target behaviors loaded cleanly onto two factors (see Table 12)[8]: (1) behaviors that are relatively mild (picketing, staging a sit-in) and (2) behaviors that are relatively severe (pushing targets lightly, pushing targets hard, kicking targets, spitting at targets, using pepper spray to temporarily blind targets, throwing small objects at targets). Based on these results, I ran separate models below within both studies: one for each factor underlying the dependent measures.

---

[8] This two-factor solution emerged regardless of whether I averaged across the three dependent measures or analyzed them separately.

Table 12. *Principal Components Analysis of Protest Behaviors in Studies 1 & 2 (Combined Sample)*

	Item	Factor 1: Severe Harms	Factor 2: Mild Harms
Behavior 1	Holding picket signs to show opposition to obnoxious [target]	0.02	**0.95**
Behavior 2	Staging a sit-in in a public place to send a message to obnoxious [target]	0.08	**0.94**
Behavior 3	Pushing obnoxious [target] hard	**0.94**	0.06
Behavior 4	Throwing small objects at obnoxious [target]	**0.94**	0.01
Behavior 5	Kicking obnoxious [target]	**0.96**	-0.01
Behavior 6	Spitting at obnoxious [target]	**0.92**	0.00
Behavior 7	Using pepper spray to temporarily blind obnoxious [target]	**0.85**	0.08
Behavior 8	Pushing obnoxious [target] lightly	**0.88**	0.15
	Eigenvalue	5.06	1.78
	% variance explained	63.25%	22.19%

Preliminary Results

Bot Detection

Eight hundred fifty-nine participants passed the Captcha question. Fourteen participants failed the Captcha question, all of whom were excluded from the analyses below.

Attention Checks

I administered the same attention check measures as the Pilot studies.

Participants self-reported on 5-point scales that I should *very much* trust their data ($M = 4.81$, $SD = 0.52$), that they were *not at all* distracted while completing the study ($M = 1.22$, $SD = 0.65$), and that they were *very much* paying attention during the study ($M = 4.85$, $SD = 0.43$). Only four participants fell below a self-reported mean level of *moderately* attentive. Corroborating the finding that most participants were attentive, 79.8% of participants correctly

identified their favorite movie for the open-ended attention check. In contrast, 8.9% of participants incorrectly selected "multiculturalism," 0.2% of participants incorrectly selected "sexism," and 11.0% skipped the question.

Manipulation Check

I administered one manipulation check item at the end of the survey: "In this study you were asked to report the extent to which you would support/oppose enacting various behaviors toward protesters at upcoming Alt-Right/Antifa rally. What protesters did you evaluate during this task?" Participants had the answer choices *supporters of multiculturalism*, *opponents of multiculturalism*, or *another group not listed here*. Participants passed the manipulation check if they correctly identified the targets to which they were randomly assigned to evaluate. Surprisingly, only 75.9% of participants correctly identified their assigned targets. Although a tentative possibility, a quarter of the sample may have failed the manipulation check because when evaluating their assigned targets, they may have also frequently considered members of the opposing side. This could have led to confusion about what the manipulation check item was asking, if they did indeed consider members of both sides while completing the survey. Further quelling fears of extremely inattentive participants, I repeatedly reminded participants of their assigned targets (in total, 24 times across the items that comprise the dependent measures). For these reasons, I decided to treat whether participants passed/failed the manipulation check as a moderator of my primary analyses rather than an automatic exclusion criterion (for more details, see "Exclusion Decisions" below).

Exclusion Decisions

For simplicity, I computed one categorical variable that reflected whether participants passed *all* attention checks and the manipulation check ($N = 531$) or failed at least one of these

measures ($N = 247$). Note that for the three continuous self-reported attention check items, I counted anyone who self-reported that they were less than a mean of *moderately* attentive as a "fail."

I ran the results described below both including and excluding participants who failed at least one attention measure. Core findings did not vary as a function of this decision rule for either of the main studies, so I ultimately retained them for the analyses described below.[9]

Study 1 Results

The purpose of Study 1 was to explore participants' willingness to harm obnoxious opponents of multiculturalism. From my total sample comprised of both supporters and opponents of multiculturalism, 397 of them were randomly assigned to evaluate *opponents* of multiculturalism and were therefore included in the Study 1 sub-sample (see Study 2 for those who were randomly assigned to evaluate *supporters* of multiculturalism; $N = 396$).

Primary Analyses

See Table 13 for the correlations between all variables in Study 1, as well as descriptive statistics for these variables. Upon inspecting these correlations, it is interesting to note that moral conviction was not associated with any harm endorsement measure, perhaps because these correlations encompass evaluations of obnoxious opponents of multiculturalism from the perspectives of both supporters and opponents of multiculturalism. That is, it may be possible that the link between moral conviction and harm endorsement is moderated by multiculturalism attitude congruence with the target, a possibility I will explore in my analyses below.

---

[9] Study 1: Core findings that link moral conviction about multiculturalism with endorsement of relatively mild behavior emerge regardless of exclusion decisions. Study 2: Core findings that link stronger moral conviction about multiculturalism with weaker endorsement non-moral justifications, and in turn, weaker endorsement of severe harms emerge regardless of exclusion decisions.

Furthermore, moral conviction was *positively* associated with ethical idealism, that is, the belief that it is never acceptable to sacrifice others' welfare to serve a cause. My hypothesizing about how moral convictions license harming opponents in the context of intergroup conflict predicts the opposite: that stronger moral convictions should weaken endorsement of ethical idealism. I elaborate on this unexpected finding below and explore whether it is moderated by multiculturalism attitude congruence (see Testing the Relationship Between Moral Conviction and Ethical Idealism).

Interestingly, non-moral justifications were consistently positively correlated with harm endorsement and moral self-image, but negatively correlated with ethical idealism. Greater endorsement of advantageous comparison, euphemistic labeling, denial of consequences, and dehumanization was consistently associated with greater harm endorsement, greater moral self-image ratings, and lower endorsement of ethical idealism.

Finally, greater liberalism and support for multiculturalism were associated with greater harm endorsement, a finding that makes sense given that the target of evaluation in Study 1 was obnoxious *opponents* of multiculturalism (a position that was also spontaneously associated with perceived conservatism among Pilot study respondents); in other words, people were more likely to endorse harm the more oppositional they were to their assigned target.

Table 13. *Correlations and Descriptive Statistics for all Study 1 Variables. Target = Opponents of Multiculturalism.*

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. MC	3.97	0.99	--	0.06	-0.12*	-0.09	0.11*	0.23***	-0.04	-0.01	0.03	0.11*	0.26***	-0.31***	-0.20***
2. Advantageous Comp	2.49	1.19		--	0.26***	0.08	0.32***	0.23***	0.08	0.13**	0.10*	0.22***	-0.07	0.09	0.01
3. Euphemistic Label	1.44	0.81			--	0.57***	0.29***	0.16**	0.29***	0.36***	0.29***	0.22***	-0.35***	0.13*	-0.05
4. Denial of Consequences	1.54	0.86				--	0.13*	0.09	0.25***	0.30***	0.25***	0.09	-0.38***	0.07	-0.08
5. Dehumanization	2.38	1.26					--	0.47***	0.17**	0.20***	0.15**	0.27***	-0.10*	-0.11*	-0.07
6. Threat	2.56	1.12						--	0.06	0.08	0.08	0.17**	-0.04	-0.08	-0.11*
7. Harm Endorsement: Support/Oppose	-1.44	1.09							--	0.85***	0.81***	-0.05	-0.52***	-0.12*	-0.26***
8. Harm Endorsement: Encourage/Discourage	-1.46	1.12								--	0.87***	0.04	-0.49***	-0.13*	-0.25***
9. Harm Endorsement: Justifiable/Unjustifiable	-1.28	1.13									--	-0.02	-0.50***	-0.16**	-0.26***
10. Moral Image	0.05	0.91										--	0.01	-0.05	-0.06
11. Ethical Idealism	1.66	1.16											--	-0.15**	0.13**
12. Multiculturalism attitude position	-2.08	2.34												--	0.47***
13. Political orientation	0.18	2.71													--

Note. * *p* < 0.05; ** *p* < .01; *** *p* < .001. All variables are measured on 1-5 scales, except the three harm endorsement scales (-3 to +3), moral image (-3 to +3), ethical idealism (-3 to +3), attitudes toward multiculturalism (-4 *very much support* to +4 *very much oppose*) and political orientation (-4 *very much liberal* to +4 *very much conservative*). Greater scores on variables generally reflect greater endorsement of the constructs, except for multiculturalism attitudes and political orientation. For more detail about how variables were measured, see Method section of Study 1.

Controlling for Attitude Strength

The zero-order correlation between moral conviction about multiculturalism and attitude strength for that attitude object was very high, $r = 0.794$. Due to concerns about multicollinearity, I did not control for attitude strength in the analyses below.

Moderated Mediation: Moral Conviction x Attitude Congruence → Harm Endorsement → Moral Self-Image

The *moral motivation and incongruence hypotheses* predict that people will be more willing to harm oppositional (vs. likeminded) others in an effort to achieve strongly (vs. weakly) moralized ends, which in turn should boost moral self-image because it is construed as a morally courageous act that serves a higher moral purpose. If the moral motivation and incongruence hypotheses are true, greater moral conviction about multiculturalism should predict greater endorsement of all harmful behaviors at the rally—both relatively mild and severe—toward oppositional (but not likeminded) protesters, which in turn should predict stronger moral self-image ratings.

These hypotheses suggest a partial mediation of the moderated link between moral conviction and moral self-image through mild and severe harm endorsement (see Figure 3). The direct effect of moral conviction on moral self-image should be positive and stronger among supporters of multiculturalism whose attitudes are incongruent with the target's stance (i.e., greater moral conviction should directly predict greater moral self-image), but the direct link between moral conviction and moral self-image should be non-significant or weaker among opponents of multiculturalism who are attitudinally congruent with the target. Moreover, these hypotheses suggest a significant/stronger indirect effect of moral conviction on moral self-image through harm endorsement for supporters of multiculturalism who are attitudinally incongruent

with the target (i.e., greater moral conviction → greater harm endorsement → greater moral self-image), but a non-significant or weaker indirect effect among opponents of multiculturalism who agree with the target's stance. These hypotheses were largely unsupported.

To test the hypothesis that stronger moral conviction predicts stronger moral self-image through harm endorsement, I used Hayes' (2018) PROCESS macro Model 8 and 5,000 bootstrapped resamplings. I entered moral conviction about multiculturalism as the focal predictor, endorsement of relatively mild and severe protest behaviors as mediators, moral self-image as the dependent variable, and attitudinal incongruence with the target (0 = congruence/opponents of multiculturalism; 1 = incongruence/supporters of multiculturalism) as a moderator of the A paths and the direct path.

Direct effects. I first examined the direct effect of moral conviction on moral self-image, and whether the direct effect was moderated by multiculturalism attitude congruence. The hypothesized conditional direct effect was non-significant as were all other direct effect estimates (see Table 14 or Table 15 for all direct effect estimates, and most notably, the null moral conviction x attitude congruence direct effect).

Indirect effects through mild harms. Next, I examined the indirect effect of moral conviction on moral self-image through mild harm endorsement at each level of multiculturalism attitude congruence (see Table 14 for all path estimates through mild harm endorsement[10]). Results revealed that, although greater moral conviction significantly predicted greater mild harm endorsement for respondents who were attitudinally incongruent (but not congruent) with

---

[10] I report each mediator in a separate table for simplicity, but I entered both mediators into one overall moderated mediation model. See Table 15 for all path estimates through severe harm endorsement.

the target as hypothesized[11], mild harm endorsement in turn unexpectedly predicted significantly

lower levels of moral self-image. That is, greater moral conviction predicted greater endorsement

of mild protest behaviors when the target was oppositional (but not likeminded), which in turn

predicted lower levels of moral self-image (see Figure 6). Corroborating this result, the index of

moderated mediation was significant through mild harm endorsement, index = -0.06, SE_{boot} =

0.03, 95% CI: [-0.13, -0.01].

Table 14. *Moral Conviction → Mild Harm Endorsement → Moral Self Image: All Path Estimates*

Predictor	Mild Harm Endorse (M)		Moral Self-Image (Y)	
Moral conviction (X)	$a_1 \rightarrow$	-0.03 (0.17), [-0.37, 0.31]	$c'_1 \rightarrow$	0.09 (0.10), [-0.10, 0.28]
Participant multiculturalism attitude (W) 0 = Oppose, 1 = Support	$a_2 \rightarrow$	-1.30 (0.78), [-2.83, 0.22]	$c'_2 \rightarrow$	-0.18 (0.42), [-1.01, 0.66]
Mild harm endorsement (M)			$b \rightarrow$	**-0.11 (0.03), [-0.16, -0.05]****
Moral conviction x Multiculturalism attitude (X x W)	$a_3 \rightarrow$	**0.60 (0.20), [0.21, 0.99]****	$c'_3 \rightarrow$	0.08 (0.11), [-0.13, 0.30]

Note. ***p < 0.001; **p < 0.01; *p < 0.05. Path coefficients are unstandardized, with standard errors represented in parentheses and 95% confidence intervals in brackets.

---

[11] Unpacking this interaction revealed that, as expected, greater moral conviction predicted greater endorsement of mild behaviors when the target was *incongruent* with participants' stances on the issue/among participants who supported multiculturalism, B = 0.57, SE = 0.10, $t(390)$ = 5.98, *p* < .001, but moral conviction did not predict endorsement of mild behaviors when the target was *congruent* with participants' beliefs/when participants opposed multiculturalism, B = -0.03, SE = 0.17, $t(390)$ = -0.17, *p* = 0.87.
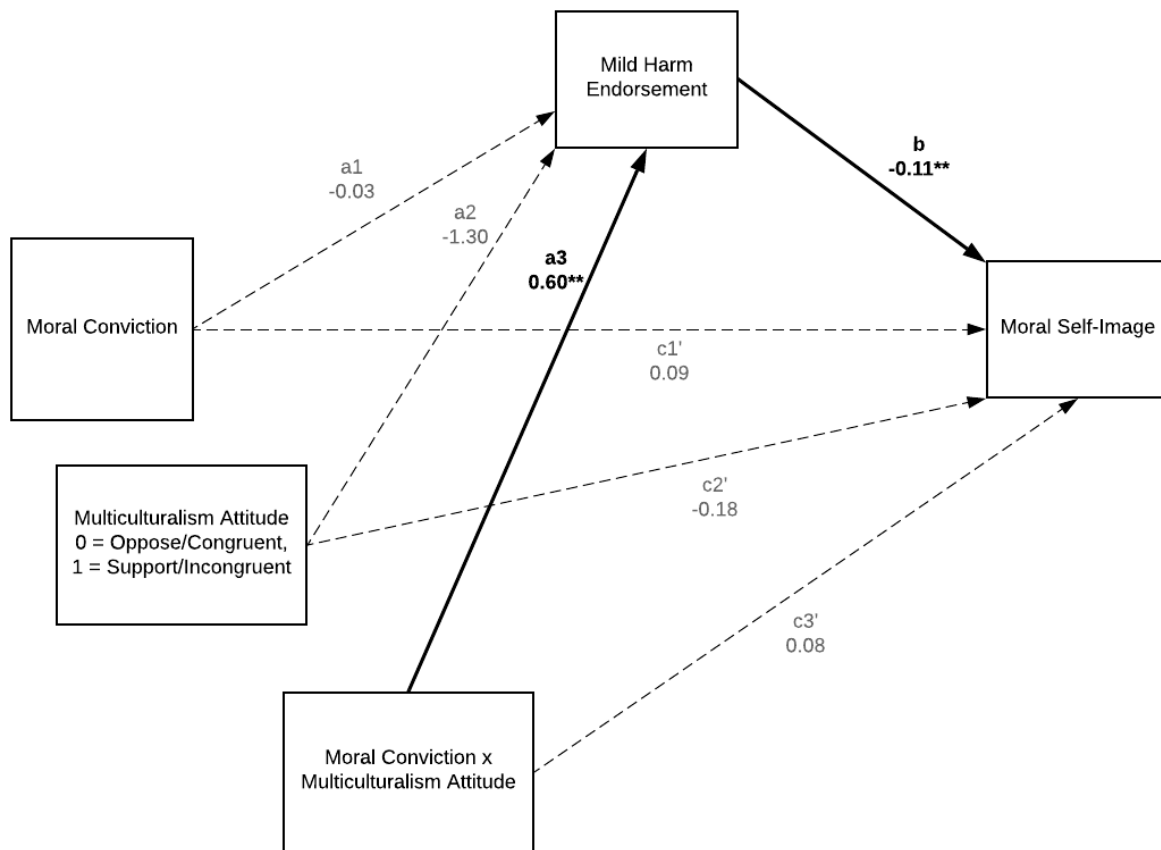
Figure 6. Statistical diagram of all path estimates represented in Table 14. Bolded paths are significant.

Indirect effects through severe harm endorsement. Next, I inspected the conditional indirect effect of moral conviction on moral self-image through severe harms (see Table 15 and Figure 7). Results revealed that stronger moral conviction unexpectedly predicted *weaker* harm endorsement toward obnoxious opponents of multiculturalism, an effect that was unmoderated by multiculturalism attitude congruence. Moreover, severe harm endorsement unexpectedly *did not* predict moral self-image. In sum, there was no evidence of a conditional indirect effect of moral conviction on moral self-image through severe harm endorsement, as confirmed by the null index of moderated mediation, index = 0.01, SE_{boot} = 0.01, 95% CI: [-0.01, 0.04].

Table 15. *Moral Conviction → Severe Harm Endorsement → Moral Self Image: All Path Estimates*

Predictor		Severe Harm Endorse (M)		Moral Self-Image (Y)
Moral conviction (X)	$a_1 \rightarrow$	**-0.33 (0.13), [-0.58, -0.07]***	$c'_1 \rightarrow$	0.09 (0.10), [-0.10, 0.28]
Participant multiculturalism attitude (W) 0 = Oppose, 1 = Support	$a_2 \rightarrow$	-0.54 (0.58), [-1.69, 0.60]	$c'_2 \rightarrow$	-0.18 (0.42), [-1.01, 0.66]
Severe harm endorsement (M)			$b \rightarrow$	0.06 (0.04), [-0.02, 0.13]
Moral conviction x Multiculturalism attitude (X x W)	$a_3 \rightarrow$	0.20 (0.15), [-0.09, 0.49]	$c'_3 \rightarrow$	0.08 (0.11), [-0.13, 0.30]

Note. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. Path coefficients are unstandardized, with standard errors represented in parentheses and 95% confidence intervals in brackets.
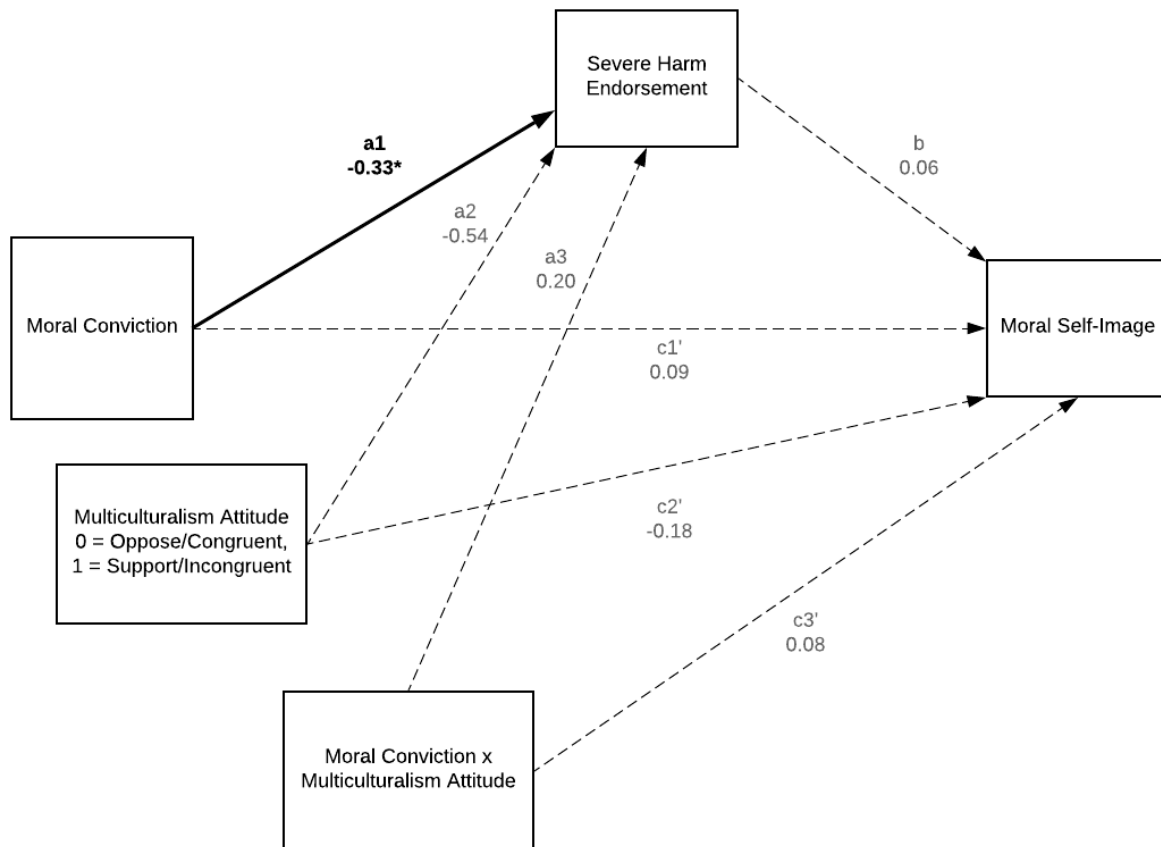

Figure 7. Statistical diagram of all path estimates represented in Table 15. Bolded paths are significant.

Summary. Overall, greater moral conviction predicted greater endorsement of mild protest behaviors when the target was oppositional (but not likeminded), which in turn predicted lower levels of moral self-image. In contrast, although greater moral conviction predicted weaker endorsement of severe protest behaviors regardless of multiculturalism attitude congruence, severe harm endorsement did not go on to predict moral self-image. Together, these results provide limited support for the moral motivation hypothesis (i.e., greater moral conviction → greater harm endorsement → greater moral self-image) and limited support for the incongruence hypothesis (i.e., that the mediational effect described by the moral motivation hypothesis is moderated by multiculturalism attitude congruence).

Moderated Mediation: Moral Conviction x Attitude Congruence → Non-Moral Justifications Endorsement → Harm Endorsement

The *moral primacy and incongruence hypotheses* state that stronger moral conviction should predict greater harm endorsement (both relatively mild and severe) against oppositional (but not likeminded) targets through non-moral justifications as mediators (i.e., advantageous comparison, euphemistic labeling, denial of consequences, and dehumanization). That is, moral conviction should serve as the primary predictor of harm endorsement toward opponents (but not likeminded targets); non-moral justifications should serve only as secondary drivers of harm endorsement when strong moral motivation is present (not when there is no moral motivation).

In other words, these hypotheses suggest a partial, conditional mediation of the link between moral conviction and harm endorsement through non-moral justifications (see Figure 4). More specifically, the direct effect of moral conviction on harm endorsement (both mild and severe) should be positive and stronger for supporters of multiculturalism who are attitudinally incongruent with the target (i.e., greater moral conviction should predict greater harm

endorsement). In contrast, the direct effect should be non-significant or weaker for opponents of multiculturalism who are attitudinally congruent with the target. Moreover, these hypotheses suggest that the indirect effect of moral conviction on harm endorsement through non-moral justification endorsement should be significant and stronger for those who are attitudinally incongruent with the target (i.e., greater moral conviction → greater non-moral justification endorsement → greater harm endorsement), but non-significant or weaker among those who are attitudinally congruent with the target. Finally, if moral conviction is the primary driver of harm endorsement, then competing moderated mediation models placing each non-moral justification as the focal predictor rather than moral conviction should be non-significant (or a worse model fit compared to the hypothesized model). These hypotheses were unsupported.

Once again, I used Hayes' (2018) PROCESS macro Model Template 8 with 5,000 bootstrapped resamplings to estimate two models: one for each kind of harm endorsement. These models included moral conviction as the focal predictor; endorsement of advantageous comparison, euphemistic labeling, denial of consequences, and dehumanization as mediators; endorsement of protest behaviors (relatively mild/severe) as the dependent variables; and participants' multiculturalism attitudes as a moderator of the A paths and the direct path (0 = attitudinal congruence with target/opponents of multiculturalism; 1 = attitudinal incongruence with target/supporters of multiculturalism).

Direct effects.

Mild harm endorsement. I first examined the conditional direct effect of moral conviction on mild harm endorsement. As hypothesized, the conditional direct effect was significant (see Table 16, particularly the moral conviction x attitude congruence direct effect estimate). Specifically, the direct effect of moral conviction on mild harm endorsement was

significant when participants supported multiculturalism/when they were incongruent with the target's stance, such that greater moral conviction predicted greater endorsement of mild harms, B = 0.52, SE = 0.09, $t(390)$ = 5.53, p < 0.001. In contrast, the direct effect of moral conviction on mild harm endorsement was not significant when participants opposed multiculturalism/were congruent with the target's stance, B = -0.09, SE = 0.17, $t(389)$ = -0.50, p = 0.62. In short, stronger moral conviction about multiculturalism predicted greater endorsement of mild harms, but only among participants who supported multiculturalism/were attitudinally incongruent with the target (not those who opposed the issue/were attitudinally congruent).

Severe harm endorsement. Next, I examined the conditional direct effect of moral conviction on <u>severe</u> harm endorsement. Contrary to predictions, the conditional direct effect of moral conviction on severe harm endorsement was non-significant. However, there was a direct effect of moral conviction on severe harm endorsement that was unmoderated by attitude congruence, such that greater moral conviction directly predicted *weaker* severe harm endorsement (see Table 17 for all direct effect estimates).

Indirect effects.

Moral conviction → Non-moral justification endorsement → Mild harm endorsement. I next examined the indirect effect of moral conviction on mild harm endorsement through non-moral justifications, and whether those paths were moderated by multiculturalism attitude congruence (see Table 16 and Figure 8 for all path estimates). Contrary to hypotheses, the indirect effects of moral conviction on harm endorsement through non-moral justifications were all non-significant, as were all indices of moderated mediation[12]. Interestingly, even though

---

[12] Advantageous comparison: index = -0.01, SE_{boot} = 0.02, 95% CI: [-0.06, 0.03]; euphemistic labeling: index = 0.01, SE_{boot} = 0.05, 95% CI: [-0.09, 0.09]; denial of consequences: index = -

moral conviction did not predict euphemistic labeling endorsement, greater euphemistic labeling predicted weaker and not stronger endorsement of mild harms.

In short, there was no evidence that the link between moral conviction and mild harm endorsement was mediated by non-moral justifications. Because the moral primacy hypothesis was not at all supported, I did not analyze competing moderated mediation models, which would move each non-moral justification to the focal predictor position in turn to confirm those alternative models fit the data less well than the current model with moral conviction as the focal predictor.

0.01, $SE_{boot} = 0.02$, 95% CI: [-0.05, 0.02]; and dehumanization: index = 0.00, $SE_{boot} = 0.01$, 95% CI: [-0.03, 0.03].

Table 16. *Moral Conviction* → *Non-Moral Justification Endorsement* → *Mild Harm Endorsement: All Path Estimates*

Predictor		Advantageous Comp (*M1*)		Mild Harm Endorsement (*Y*)
Moral conviction (*X*)	a_{1Comp}→	0.15 (0.13), [-0.10, 0.40]	c'_1 →	-0.09 (0.17), [-0.42, 0.25]
Participant multiculturalism attitude (*W*) 0 = Oppose, 1 = Support	a_{2Comp}→	-0.02 (0.56), [-1.13, 1.08]	c'_2 →	-1.34 (0.76), [-2.83, 0.14]
Advantageous Comp (*M1)*			b_{Comp} →	0.11 (0.07), [-0.03, 0.26]
Moral conviction x Multiculturalism attitude (*X* x *W*)	a_{3Comp}→	-0.08 (0.14), [-0.36, 0.21]	c'_3 →	**0.61 (0.19), [0.23, 0.99]****
		Euphemistic Labeling (*M2*)		
Moral conviction (*X*)	a_{1Euph}→	-0.08 (0.09), [-0.25, 0.09]	c'_1 →	-0.09 (0.17), [-0.42, 0.25]
Participant multiculturalism attitude (*W*) 0 = Oppose, 1 = Support	$a_{2\ Euph}$ →	-0.07 (0.38), [-0.82, 0.68]	c'_2 →	-1.34 (0.76), [-2.83, 0.14]
Euphemistic Labeling (*M2*)			b_{Euph} →	**-0.41 (0.13), [-0.66, -0.16]****
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\ Euph}$ →	-0.02 (0.10), [-0.21, 0.18]	c'_3 →	**0.61 (0.19), [0.23, 0.99]****

Denial of Consequences (*M3*)

Moral conviction (*X*)	$a_{1Deny} \rightarrow$	-0.12 (0.09), [-0.30, 0.06]	$c'_1 \rightarrow$	-0.09 (0.17), [-0.42, 0.25]
Participant multiculturalism attitude (*W*) 0 = Oppose, 1 = Support	$a_{2\,Deny} \rightarrow$	-0.23 (0.41), [-1.03, 0.58]	$c'_2 \rightarrow$	-1.34 (0.76), [-2.83, 0.14]
Denial of Consequences (*M3*)			$b_{Deny} \rightarrow$	-0.11 (0.11), [-0.33, 0.12]
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\,Deny} \rightarrow$	0.06 (0.11), [-0.15, 0.26]	$c'_3 \rightarrow$	**0.61 (0.19), [0.23, 0.99]****

Dehumanization (*M4*)

Moral conviction (*X*)	$a_{1Dehum} \rightarrow$	0.12 (0.13), [-0.15, 0.38]	$c'_1 \rightarrow$	-0.09 (0.17), [-0.42, 0.25]
Participant multiculturalism attitude (*W*) 0 = Oppose, 1 = Support	$a_{2\,Dehum} \rightarrow$	0.18 (0.60), [-1.00, 1.35]	$c'_2 \rightarrow$	-1.34 (0.76), [-2.83, 0.14]
Dehumanization (*M4*)			$b_{Dehum} \rightarrow$	-0.06 (0.07), [-0.19, 0.08]
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\,Dehum} \rightarrow$	0.01 (0.15), [-0.29, 0.31]	$c'_3 \rightarrow$	**0.61 (0.19), [0.23, 0.99]****

Note. ***p < 0.001; **p < 0.01; *p < 0.05. Path coefficients are unstandardized, with standard errors represented in parentheses and 95% confidence intervals in brackets.
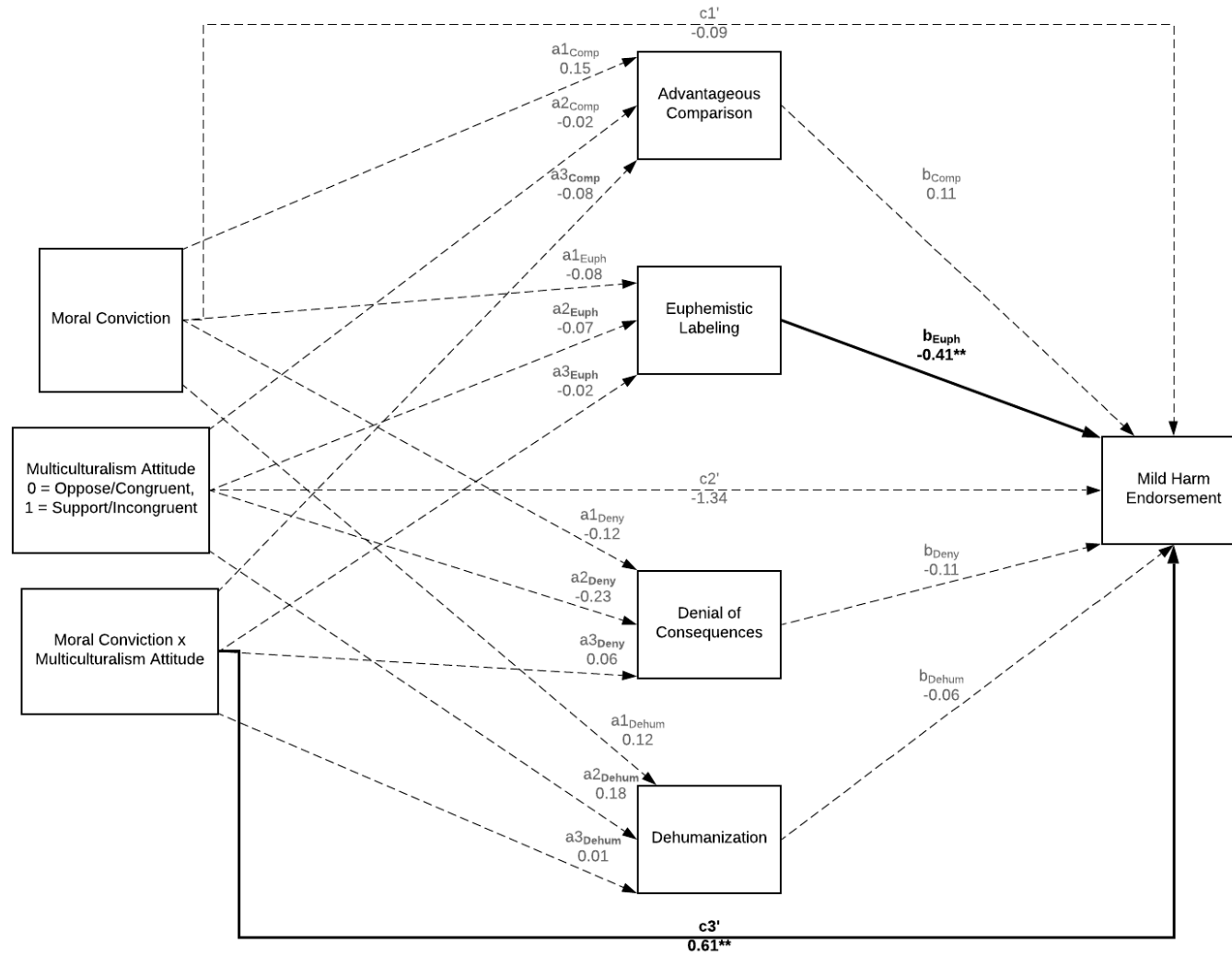
Figure 8. Statistical diagram of all path estimates represented in Table 16. Bolded paths are significant.

Moral conviction → Non-moral justification endorsement → Severe harm

endorsement. Then, I inspected the indirect effects of moral conviction on severe harm

endorsement through non-moral justifications and whether multiculturalism attitude congruence

moderated these effects (see Table 17 and Figure 9 for all path estimates).

Once again, the indirect effects of moral conviction on severe harm endorsement through non-

moral justifications were all non-significant, as were the indices of moderated mediation.[13]

Interestingly, even though moral conviction did not predict any non-moral justification, greater

endorsement of euphemistic labeling, denial of consequences, and dehumanization predicted

greater endorsement of severe harms.

Overall, the moral primacy and incongruence hypotheses were not supported, and I

therefore did not explore competing moderated mediation models.

---

[13] Advantageous comparison: index = 0.00, SE_{boot} = 0.01, 95% CI: [-0.02, 0.02]; euphemistic labeling: index = -0.01, SE_{boot} = 0.06, 95% CI: [-0.12, 0.11]; denial of consequences: index = 0.01, SE_{boot} = 0.03, 95% CI: [-0.04, 0.07]; and dehumanization: index = 0.00, SE_{boot} = 0.02, 95% CI: [-0.04, 0.04].

Table 17. *Moral Conviction → Non-Moral Justification Endorsement → Severe Harm Endorsement: All Path Estimates*

Predictor		Advantageous Comp (*M1*)		Severe Harm Endorsement (*Y*)
Moral conviction (*X*)	a_{1Comp}→	0.15 (0.13), [-0.10, 0.40]	c'_1 →	**-0.26 (0.11), [-0.49, -0.04]***
Participant multiculturalism attitude (*W*) 0 = Oppose, 1 = Support	a_{2Comp}→	-0.02 (0.56), [-1.13, 1.08]	c'_2 →	-0.46 (0.50), [-1.45, 0.52]
Advantageous Comp (*M1*)			b_{Comp} →	0.00 (0.05), [-0.09, 0.10]
Moral conviction x Multiculturalism attitude (*X* x *W*)	a_{3Comp}→	-0.08 (0.14), [-0.36, 0.21]	c'_3 →	0.19 (0.13), [-0.06, 0.45]
Euphemistic Labeling (*M2*)				
Moral conviction (*X*)	a_{1Euph}→	-0.08 (0.09), [-0.25, 0.09]	c'_1 →	**-0.26 (0.11), [-0.49, -0.04]***
Participant multiculturalism attitude (*W*) 0 = Oppose, 1 = Support	$a_{2\ Euph}$ →	-0.07 (0.38), [-0.82, 0.68]	c'_2 →	-0.46 (0.50), [-1.45, 0.52]
Euphemistic Labeling (*M2*)			b_{Euph} →	**0.54 (0.09), [0.37, 0.70]*****
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\ Euph}$ →	-0.02 (0.10), [-0.21, 0.18]	c'_3 →	0.19 (0.13), [-0.06, 0.45]

Denial of Consequences (*M3*)

Moral conviction (*X*)	$a_{1Deny} \rightarrow$	-0.12 (0.09), [-0.30, 0.06]	$c'_1 \rightarrow$	**-0.26 (0.11), [-0.49, -0.04]***
Participant multiculturalism attitude (*W*) 0 = Oppose, 1 = Support	$a_{2\,Deny} \rightarrow$	-0.23 (0.41), [-1.03, 0.58]	$c'_2 \rightarrow$	-0.46 (0.50), [-1.45, 0.52]
Denial of Consequences (*M3*)			$b_{\,Deny} \rightarrow$	**0.26 (0.07), [0.11, 0.41]****
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\,Deny} \rightarrow$	0.06 (0.11), [-0.15, 0.26]	$c'_3 \rightarrow$	0.19 (0.13), [-0.06, 0.45]

Dehumanization (*M4*)

Moral conviction (*X*)	$a_{1Dehum} \rightarrow$	0.12 (0.13), [-0.15, 0.38]	$c'_1 \rightarrow$	**-0.26 (0.11), [-0.49, -0.04]***
Participant multiculturalism attitude (*W*) 0 = Oppose, 1 = Support	$a_{2\,Dehum} \rightarrow$	0.18 (0.60), [-1.00, 1.35]	$c'_2 \rightarrow$	-0.46 (0.50), [-1.45, 0.52]
Dehumanization (*M4*)			$b_{\,Dehum} \rightarrow$	**0.10 (0.05), [0.01, 0.19]***
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\,Dehum} \rightarrow$	0.01 (0.15), [-0.29, 0.31]	$c'_3 \rightarrow$	0.19 (0.13), [-0.06, 0.45]

Note. ***p < 0.001; **p < 0.01; *p < 0.05. Path coefficients are unstandardized, with standard errors represented in parentheses and 95% confidence intervals in brackets.
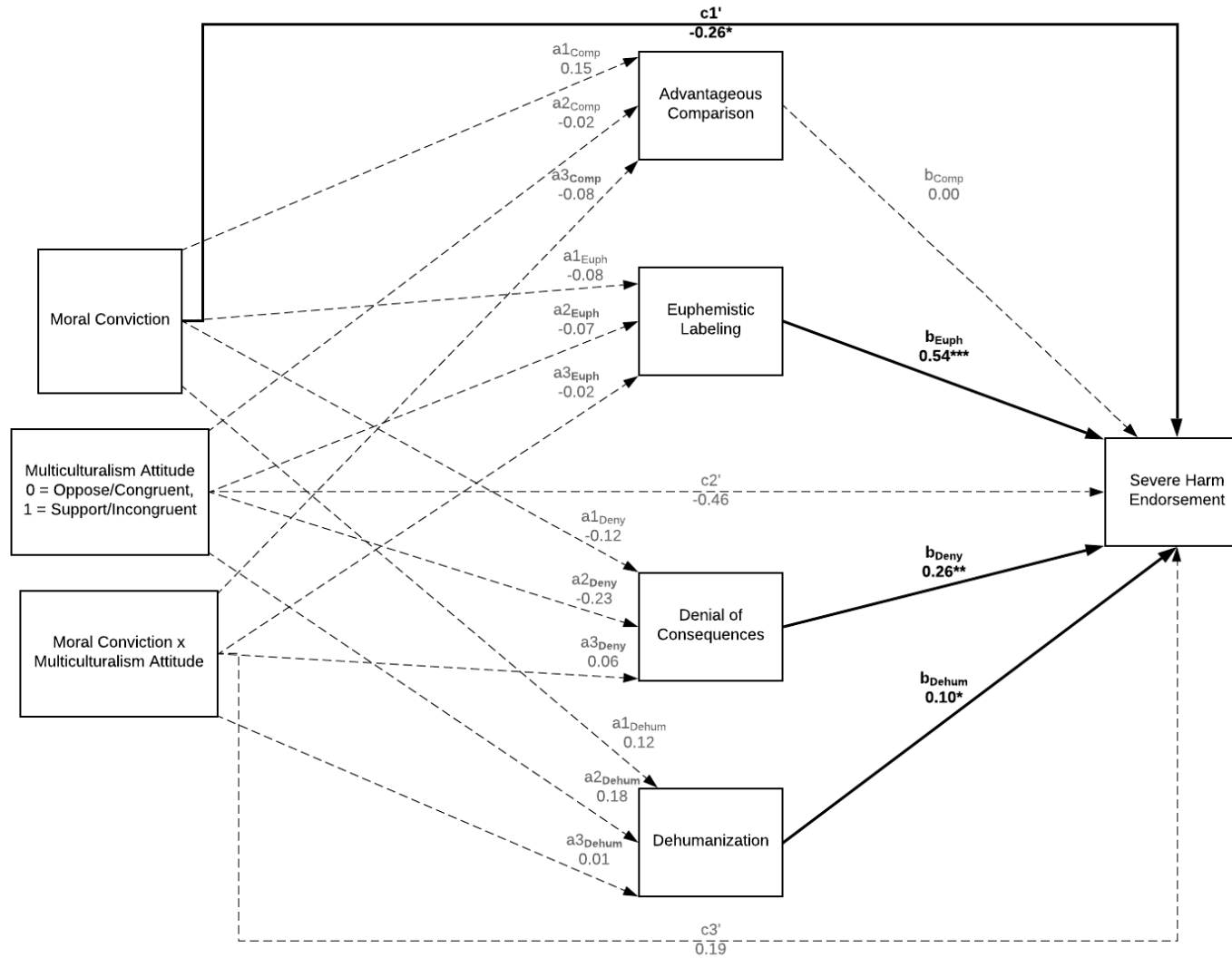
Figure 9. Statistical diagram of all path estimates represented in Table 17. Bolded paths are significant.

Summary. As hypothesized, greater moral conviction directly predicted greater endorsement of <u>mild</u> protest behaviors when the target was oppositional, but not likeminded. Unexpectedly, however, greater moral conviction directly predicted weaker endorsement of <u>severe</u> protest behaviors regardless of multiculturalism attitude congruence. Further contradicting hypotheses, none of these direct effects were mediated by non-moral justifications. Overall, these results do not provide support for the moral primacy and incongruence hypotheses, which argue that 1) moral conviction should be the primary path toward endorsing all harmful behaviors against opponents (but not likeminded targets); and 2) that non-moral justifications should mediate the link between moral conviction and harm endorsement.

Testing the Relationship Between Moral Conviction & Ethical Idealism

Although my analyses above do not support my theory of moral motivation and harm endorsement, I nonetheless explored the relationship between moral conviction and beliefs about ethical idealism (i.e., that it is never acceptable to sacrifice others' wellbeing for a noble cause). If my theorizing is correct, that strong moral mandates license harming one's opponents, then moral conviction about multiculturalism should be negatively correlated with ethical idealism.

Contrary to hypotheses, greater moral conviction about multiculturalism was associated with greater endorsement that it is never acceptable to sacrifice the welfare of others to serve personal beliefs about multiculturalism, $r(391) = 0.26$, $p < .001$. Moreover, the positive effect of moral conviction on ethical idealism was unmoderated by participants' side of issue, $B = -0.14$, $SE = 0.14$, $t(389) = -1.04$, $p = 0.30$. In other words, the finding that greater moral conviction predicts greater endorsement of ethical idealism was equally true of supporters and opponents of multiculturalism. This result provides no support for my theorizing about the link between moral conviction and harm endorsement.

Testing the Threat Hypothesis

Although I initially hypothesized that perceived worldview threat at the rally would intensify the link between moral conviction and endorsement of harms against oppositional targets through non-moral justifications, my lack of support for the moral primacy and incongruence hypotheses above (particularly how there were no significant indirect effects of moral conviction on harm endorsement through non-moral justifications) led me to pivot on my strategy for testing the threat hypothesis. To simplify the analysis, I dropped non-moral justifications as mediators and decided to instead test the three-way interaction of moral conviction, threat (continuously measured), and attitudinal incongruence with the target (0 = congruent; 1 = incongruent) predicting harm endorsement within the PROCESS macro using Model Template 3 and 5,000 bootstrapped resamplings.

The simplified threat hypothesis is nearly identical to the original version depicted in Figure 5, with the exception that I will not be testing for a mediation effect through non-moral justifications. If this simplified version of the threat hypothesis is true, then I would expect to see that, among people who support multiculturalism/are attitudinally incongruent with the assigned target, stronger perception of threat should more strongly intensify the hypothesized positive association between moral conviction and endorsement of protest behaviors (both relatively mild and severe). In contrast, among participants who oppose multiculturalism/are attitudinally congruent with the assigned target, moral conviction should not predict (or more weakly predict) endorsement of protest behaviors, regardless of perceived threat at the rally.

Relatively mild protest behaviors. The overall model was significant, $F(7, 386) = 10.72$, $p < 0.001$, $R^2 = 0.16$. Moreover, as predicted, the three-way interaction was significant (see Table 18 for all path estimates), suggesting that the interactive effect of threat and moral

conviction on mild harm endorsement varied as a function of whether participants supported or opposed multiculturalism. However, after unpacking the interaction, the threat hypothesis was unsupported for evaluations of relatively mild protest behaviors.

More specifically, the simple interaction of moral conviction and threat was significant among people who opposed multiculturalism/were attitudinally congruent with the target. Among opponents of multiculturalism who imbued the rally with low (-1 SD) levels of threat, moral conviction did not significantly predict endorsement of mild protest behaviors. Likewise, moral conviction did not predict mild protest behavior endorsement at average levels of threat for people who opposed multiculturalism. However, among opponents of multiculturalism who felt highly threatened by the rally (+1 SD), stronger moral conviction marginally predicted weaker endorsement of relatively mild protest behaviors.

In contrast, the simple interaction of threat and moral conviction was not significant among people who supported multiculturalism/were attitudinally incongruent with the target.

Taken together, high (vs. low or average) levels of threat appeared to marginally weaken endorsement of relatively mild protest behaviors among people who opposed multiculturalism/were attitudinally congruent with the target. In contrast, threat did not moderate the link between moral conviction and mild harm endorsement among people who supported multiculturalism/were attitudinally incongruent with the target. In short, the threat hypothesis was unsupported for endorsement of mild protest behaviors.

Relatively severe protest behaviors. The overall model was significant, $F(7, 386) = 3.62$, $p < 0.001$, $R^2 = 0.06$, and the interaction of moral conviction and threat significantly varied as a function of participants' attitudes toward multiculturalism (see Table 19). Once again,

however, further inspection of the three-way interaction revealed no support for the threat hypothesis.

Upon deconstructing the three-way interaction, the simple interaction of moral conviction and attitude incongruence was significant only among people who opposed multiculturalism/were congruent with the target's stance toward multiculturalism. For opponents of multiculturalism who viewed the rally as a low-level threat (-1 SD), moral conviction did not predict severe protest behavior endorsement. However, for opponents of multiculturalism at average levels of threat, greater moral conviction significantly predicted less endorsement of severe protest behaviors. This pattern was even more exaggerated among opponents of multiculturalism who viewed the rally as highly threatening (+1 SD): stronger moral conviction significantly predicted less severe protest behavior endorsement.

In contrast, the simple interaction of moral conviction and threat was not significant among people who supported multiculturalism/were attitudinally incongruent with the target.

Overall, the threat hypothesis was not supported. Higher levels of threat intensified the tendency for opponents of multiculturalism to less strongly endorse severe protest behaviors. Threat did not shape evaluations of severe protest behaviors for supporters of multiculturalism, however.[14]

---

[14] Based on the recommendation of a committee member, I also explored whether threat serves as a mediator in the link between moral conviction and harmful behavior endorsement, using support/opposition to multiculturalism as a moderator of this mediation model. Threat did not significantly mediate the link between moral conviction and either kind of protest behavior (mild/severe), and these results will not be explored further.

Table 18. *Moral Conviction x Threat x Attitude Congruence → Mild Harm Endorsement: All Path Estimates*

Effect	*F*	*p*	B	SE	*t*	*p*
MC x Threat x Att Congruence*			0.45	0.19	$t(386) = 2.41$	0.02
MC x Threat @ Congruent*	$F(1, 386) = 4.12$	0.04				
MC x Low (-1 SD) Threat			0.24	0.22	$t(386) = 1.11$	0.27
MC x Average Threat			-0.12	0.20	$t(386) = -0.64$	0.52
MC x High (+1 SD) Threat			-0.50	0.31	$t(386) = -1.62$	0.10
MC x Threat @ Incongruent	$F(1, 386) = 1.70$	0.20				

Note. *p < 0.05; **p < 0.01; ***p < 0.001. Significant effects are bolded.

Table 19. *Moral Conviction x Threat x Attitude Congruence → Severe Harm Endorsement: All Path Estimates*

Effect	*F*	*p*	B	SE	*t*	*p*
MC x Threat x Att Congruence*			0.28	0.14	$t(386) = 1.99$	0.047
MC x Threat @ Congruent*	$F(1, 386) = 5.92$	0.02				
MC x Low (-1 SD) Threat			-0.18	0.17	$t(386) = -1.08$	0.28
MC x Average Threat**			-0.51	0.15	$t(386) = -3.52$	< 0.01
MC x High (+1 SD) Threat**			-0.85	0.23	$t(386) = -3.69$	< 0.01
MC x Threat @ Incongruent	$F(1, 386) = 0.07$	0.79				

Note. *p < 0.05; **p < 0.01; ***p < 0.001. Significant effects are bolded.

Study 1 Discussion

Overall, I did not find compelling support for my core moral motivation and moral primacy hypotheses. Instead, I learned the following (as summarized in Table 20): (1) Greater moral conviction about multiculturalism consistently predicted greater endorsement of *mild* harms against oppositional (not likeminded) targets; (2) The positive link between moral conviction and mild harm endorsement for oppositional (not likeminded) targets went on to in turn predict <u>lower</u> moral self-image; (3) Regardless of multiculturalism attitude congruence, greater moral conviction about multiculturalism predicted <u>weaker</u> endorsement of *severe* harms; (4) higher levels of threat appeared to intensify the negative link between moral conviction and severe harm endorsement for people who were attitudinally congruent (vs. incongruent) with the target; and (5) greater euphemistic labeling, denial of consequences, and dehumanization consistently predicted greater endorsement of severe harms.

Table 20. *Study 1 Summarized Findings*

Hypothesis	Multiculturalism Attitude Congruence	Expected	Observed	Hypothesis Supported?
Moral Motivation & Incongruence: **Mild** Harms	Incongruent Attitudes	Greater MC → Greater **mild** harms → Greater moral self-image	Greater MC → Greater **mild** harms → Lower moral self-image	No
	Congruent Attitudes	Weaker links between MC, **mild** harms, moral self-image than incongruent attitudes	Non-significant links between MC, **mild** harms, moral self-image	Yes
Moral Motivation & Incongruence: **Severe** Harms	Incongruent Attitudes	Greater MC → Greater **severe** harms → Greater moral self-image	Greater MC → Weaker **severe** harms (no moderation by attitude congruence)	No
	Congruent Attitudes	Weaker links between MC, **severe** harms, moral self-image than incongruent attitudes		
Moral Primacy & Incongruence: **Mild** Harms	Incongruent Attitudes	Greater MC → Greater non-moral justifications → Greater **mild** harms	Conditional direct effect: Greater MC → Greater **mild** harms	Somewhat
	Congruent Attitudes	Weaker links between MC, non-moral justifications, and **mild** harms than incongruent attitudes	Conditional direct effect: No direct link between MC and **mild** harms	Somewhat
Moral Primacy & Incongruence: **Severe** Harms	Incongruent Attitudes	Greater MC → Greater non-moral justifications → Greater **severe** harms	Direct effect: Greater MC → Weaker **severe** harms (no moderation by attitude congruence)	No
	Congruent Attitudes	Weaker links between MC, non-moral justifications, and **severe** harms than incongruent attitudes		

	Incongruent Attitudes	Threat x MC x Attitude Congruence interaction: Higher threat should intensify the MC → Harm link	No simple interaction of threat x MC predicting harms	
Threat: **Mild** Harms				No
	Congruent Attitudes	No Threat x MC x Attitude Congruence interaction	Higher MC → Weaker harms @ high (not low or average) threat*	
	Incongruent Attitudes	Threat x MC x Attitude Congruence interaction: Higher threat should intensify the MC → Harm link	No simple interaction of threat x MC predicting harms	
Threat: **Severe** Harms				No
	Congruent Attitudes	No Threat x MC x Attitude Congruence interaction	Higher MC → Weaker harms @ high and average (not low) threat	

Note: This effect was marginally significant.

Study 2 Results

The purpose of Study 2 was to explore people's willingness to harm obnoxious supporters of multiculturalism. Like Study 1, both supporters and opponents of multiculturalism were recruited. In total, 396 participants from the total sample were randomly assigned to evaluate obnoxious *supporters* of multiculturalism and be included in this Study 2 sub-sample.

Primary Analyses

See Table 21 for the correlations between all variables in Study 2, as well as descriptive statistics for these variables. Unlike Study 1, in Study 2 moral conviction was negatively associated with all three harm endorsement variables; the stronger people's moral conviction about multiculturalism, the weaker their endorsement for harming obnoxious supporters of multiculturalism. It is important to note, however, that this relationship does not account for respondents' attitudes about multiculturalism. I will later explore whether this relationship is moderated by respondents' attitudinal congruence with the target. Similarly, providing initial evidence against the moral motivation and primacy hypotheses, moral conviction was positively associated with ethical idealism beliefs: The greater people's moral conviction about multiculturalism, the more they endorsed that it is never acceptable to sacrifice others' wellbeing to serve one's own beliefs about multiculturalism.

Moreover, all non-moral justifications were positively correlated with harm endorsement variables: The greater people's endorsement of non-moral justifications, the more they endorsed harming obnoxious supporters of multiculturalism. Corroborating these effects, there were also significant negative correlations between all non-moral justification endorsement variables and ethical idealism, such that greater non-moral justification endorsement was associated with weaker endorsement of beliefs that it is never acceptable to harm others for a noble cause.

Interestingly, greater endorsement of non-moral justifications was also associated with stronger moral self-image.

Finally, given that the target in Study 2 was obnoxious supporters of multiculturalism, it makes sense that in this study that greater opposition to multiculturalism and greater conservatism were associated with greater harm endorsement.

Table 21. *Correlations and Descriptive Statistics for all Study 2 Variables. Target = Supporters of Multiculturalism.*

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1. MC	3.98	1.02	--	-0.07	-0.17**	-0.17**	-0.18***	0.06	-0.22***	-0.19***	-0.18***	-0.05	0.25***	-0.37***	-0.36***
2. Advantageous Comp	2.44	1.16		--	0.27***	0.16**	0.35***	0.32***	0.19***	0.17**	0.15**	0.19***	-0.06	0.06	0.13**
3. Euphemistic Label	1.36	0.71			--	0.54***	0.35***	0.19***	0.29***	0.28***	0.16**	0.20***	-0.32***	0.15**	0.07
4. Denial of Consequences	1.60	0.95				--	0.21***	0.11*	0.34***	0.27***	0.18***	0.18**	-0.20***	0.12*	0.08
5. Dehumanization	2.14	1.21					--	0.37***	0.23**	0.28***	0.25***	0.26***	-0.14**	0.29***	0.28***
6. Threat	2.35	1.08						--	-0.01	0.00	-0.04	0.21***	0.02	-0.02	0.02
7. Harm Endorsement: Support/Oppose	-1.77	1.04							--	0.75***	0.73***	0.09	-0.22***	0.20***	0.17**
8. Harm Endorsement: Encourage/Discourage	-1.84	0.98								--	0.77***	0.11*	-0.28***	0.26***	0.17**
9. Harm Endorsement: Justifiable/Unjustifiable	-1.63	1.02									--	0.02	-0.25***	0.26***	0.16**
10. Moral Image	0.03	0.87										--	0.01	0.00	0.07
11. Ethical Idealism	1.60	1.13											--	-0.26***	-0.02
12. Multiculturalism attitude position	-2.15	2.30												--	0.51***
13. Political orientation	0.20	2.81													--

Note. * $p < 0.05$; ** $p < .01$; *** $p < .001$. All variables are measured on 1-5 scales, except the three harm endorsement scales (-3 to +3), moral image (-3 to +3), ethical idealism (-3 to +3), attitudes toward multiculturalism (-4 *very much support* to +4 *very much oppose*) and political orientation (-4 *very much liberal* to +4 *very much conservative*). Greater scores on variables generally reflect greater endorsement of the constructs, except for multiculturalism attitudes and political orientation. For more detail about how variables were measured, see Method section.

Controlling for Attitude Strength

Once again, the zero-order correlation between moral conviction about multiculturalism and attitude strength for that attitude object was very high, $r = 0.81$, raising concerns about multicollinearity. I therefore did not control for attitude strength in the analyses below.

Moderated Mediation: Moral Conviction x Attitude Congruence → Harm Endorsement → Moral Self-Image

Identical to Study 1, if the *moral motivation* and *incongruence* hypotheses are true, greater moral conviction about multiculturalism should predict greater endorsement of harmful behaviors at the rally toward oppositional (but not likeminded) protesters, which in turn should predict stronger moral self-image ratings.

As a reminder, these hypotheses suggest a partial mediation of the moderated link between moral conviction and moral self-image through mild and severe harm endorsement (see Figure 3). The direct effect of moral conviction on moral self-image should be more strongly positive among opponents of multiculturalism whose attitudes are incongruent with the target's stance; stronger moral conviction should directly predict greater moral self-image. In contrast, the direct link between moral conviction and moral self-image should be non-significant or weaker among supporters of multiculturalism who are attitudinally congruent with the target. Moreover, these hypotheses suggest a stronger significant indirect effect of moral conviction on moral self-image through harm endorsement for opponents of multiculturalism who are attitudinally incongruent with the target (i.e., greater moral conviction → greater harm endorsement → greater moral self-image), but a non-significant or weaker indirect effect among supporters of multiculturalism who agree with the target's stance. These hypotheses were unsupported.

To test these hypotheses, I used Hayes' (2018) PROCESS macro Model 8 and 5,000 bootstrapped resamplings. I entered moral conviction about multiculturalism as the focal predictor, endorsement of relatively mild and severe protest behaviors as mediators, moral self-image as the dependent variable, and attitudinal incongruence with the target (0 = congruence/supporters of multiculturalism; 1 = incongruence/opponents of multiculturalism) as a moderator of the A paths and the direct path.

Direct effects. I first examined the conditional direct effect of moral conviction on moral self-image. Contrary to predictions, the conditional direct effect of moral conviction on moral self-image was non-significant (see Table 22 or Table 23 for all direct effect estimates).

Indirect effects through mild harms. I then examined the indirect effect of moral conviction on moral self-image through mild harm endorsement at each level of multiculturalism attitude congruence (see Table 22 and Figure 10 for all path estimates through mild harm endorsement[15]). Contrary to the moral motivation and incongruence hypotheses, the conditional indirect effect was non-significant, as supported by a null index of moderated mediation, index = 0.00, SE_{boot} = 0.01, 95% CI: [-0.01, 0.02].

---

[15] I report each mediator in a separate table for simplicity, but I entered both mediators into one overall moderated mediation model. See Table 23 for all path estimates through severe harm endorsement.

Table 22. *Moral Conviction → Mild Harm Endorsement → Moral Self Image: All Path Estimates*

Predictor		Mild Harm Endorse (M)		Moral Self-Image (Y)
Moral conviction (X)	$a_1 \rightarrow$	-0.07 (0.11), [-0.29, 0.16]	$c'_1 \rightarrow$	-0.07 (0.05), [-0.17, 0.03]
Participant multiculturalism attitude (W) 0 = Support, 1 = Oppose	$a_2 \rightarrow$	0.51 (0.84), [-1.15, 2.16]	$c'_2 \rightarrow$	-0.57 (0.38), [-1.32, 0.19]
Mild harm endorsement (M)			$b \rightarrow$	0.00 (0.02), [-0.05, 0.04]
Moral conviction x Multiculturalism attitude (X x W)	$a_3 \rightarrow$	-0.04 (0.22), [-0.47, 0.39]	$c'_3 \rightarrow$	0.11 (0.10), [-0.09, 0.30]

Note. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. Path coefficients are unstandardized, with standard errors represented in parentheses and 95% confidence intervals in brackets.
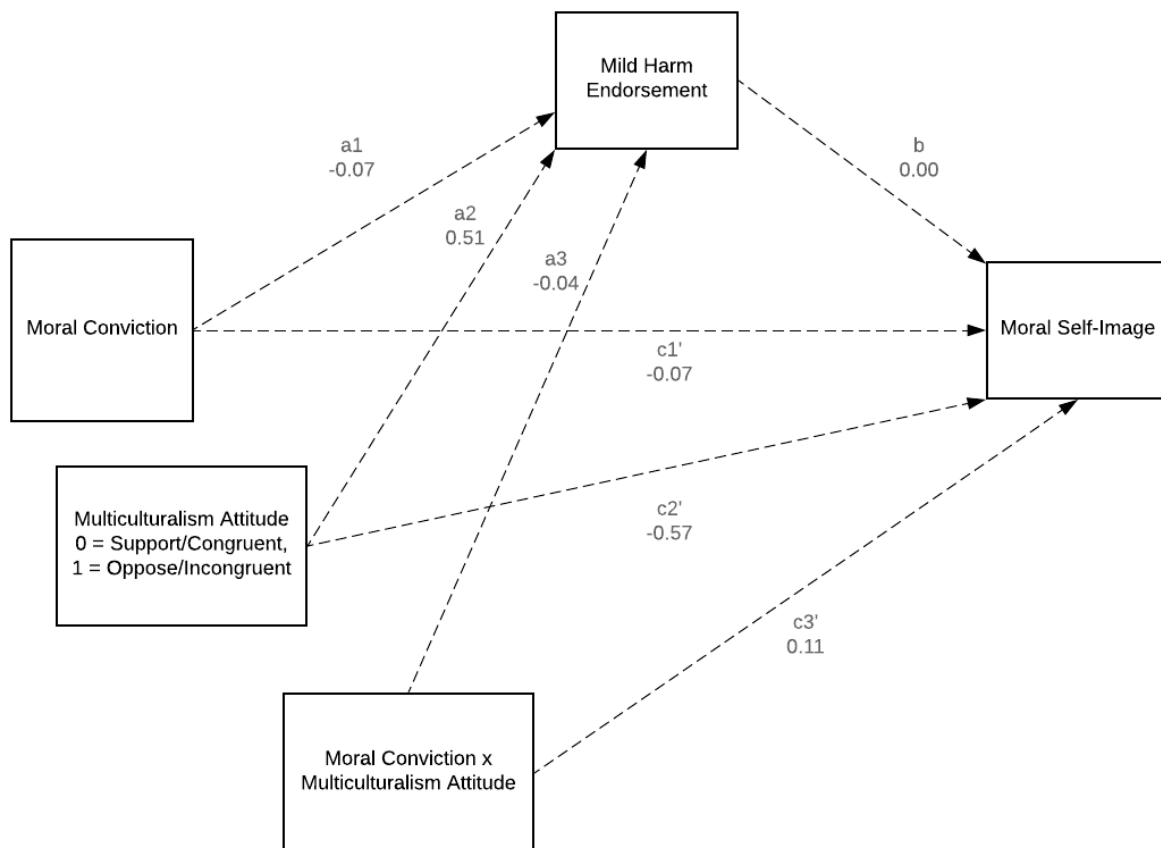

Figure 10. Statistical diagram of all path estimates represented in Table 22. All paths are non-significant.

Indirect effects through severe harms. Next, I explored the indirect effect of moral conviction on moral self-image through severe harm endorsement at each level of multiculturalism attitude congruence (see Table 23 and Figure 11 for all path estimates). Stronger moral conviction significantly predicted weaker endorsement of severe harms, which in turn predicted weaker moral self-image. However, there was no evidence that this indirect effect varied as a function of multiculturalism attitude congruence. That is, the index of moderated mediation was non-significant, index = 0.01, SE = 0.01, 95% CI: [-0.01, 0.04].[16] In simple terms, regardless of multiculturalism attitude congruence, stronger moral conviction about multiculturalism predicted weaker endorsement of severe harms, which in turn predicted weaker moral self-image.

Table 23. *Moral Conviction → Severe Harm Endorsement → Moral Self Image: All Path Estimates*

Predictor		Severe Harm Endorse (M)		Moral Self-Image (Y)
Moral conviction (X)	$a_1 →$	**-0.21 (0.06), [-0.32, -0.10]****	$c'_1 →$	-0.07 (0.05), [-0.17, 0.03]
Participant multiculturalism attitude (W) 0 = Support, 1 = Oppose	$a_2 →$	-0.01 (0.42), [-0.83, 0.81]	$c'_2 →$	-0.57 (0.38), [-1.32, 0.19]
Severe harm endorsement (M)			$b →$	**0.09 (0.05), [0.001, 0.19]***
Moral conviction x Multiculturalism attitude (X x W)	$a_3 →$	0.11 (0.11), [-0.10, 0.32]	$c'_3 →$	0.11 (0.10), [-0.09, 0.30]

Note. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. Path coefficients are unstandardized, with standard errors represented in parentheses and 95% confidence intervals in brackets.

---

[16] I confirmed that the indirect effect of moral conviction on moral self-image through severe harm endorsement was significant in a mediation model that did not explore moderation by multiculturalism attitude congruence (i.e., Model Template 4).
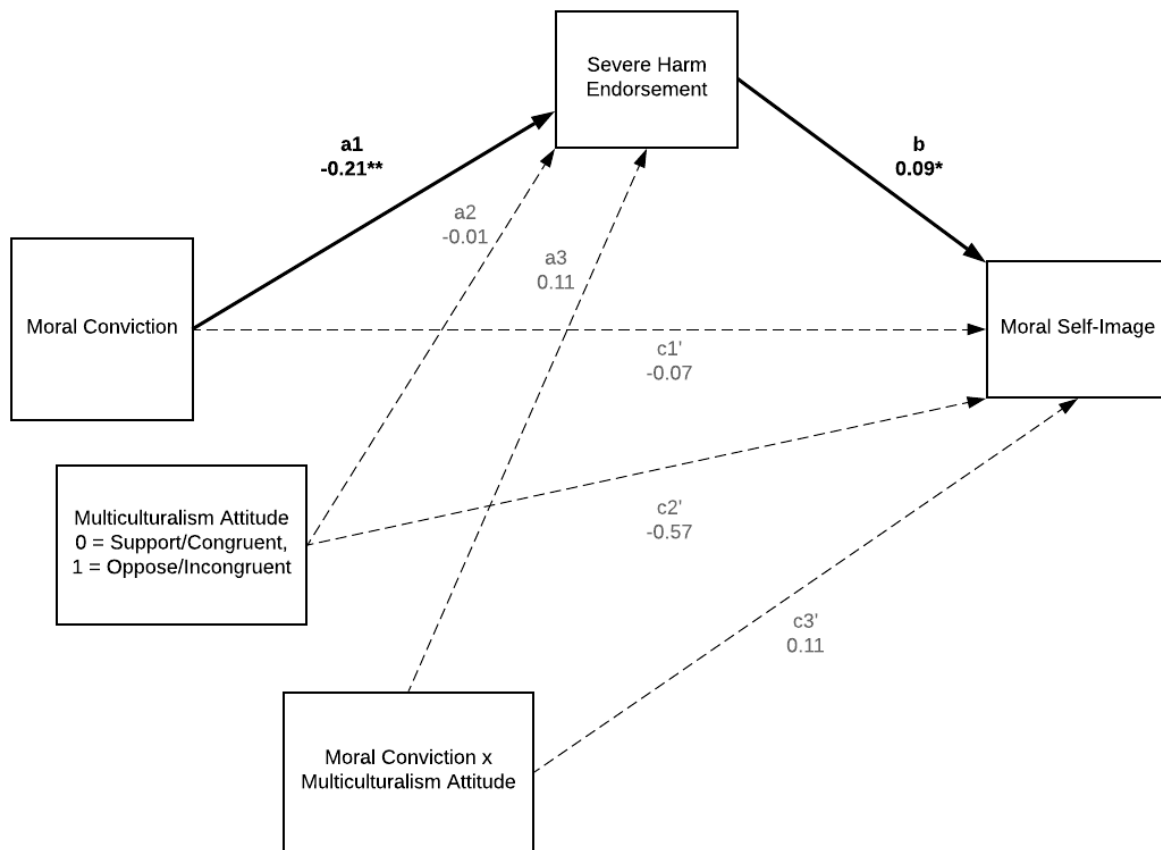
Figure 11. Statistical diagram of all path estimates represented in Table 23. Bolded paths are significant.

Summary. Contrary to hypotheses, there was no evidence that moral conviction predicted moral self-image through mild harm endorsement. However, the indirect effect of moral conviction on moral self-image was significant through severe harm endorsement (i.e., greater moral conviction → weaker severe harm endorsement → weaker moral self-image): an effect that was unmoderated by multiculturalism attitude congruence. Together, these results provide no support for the moral motivation hypothesis (i.e., greater moral conviction → greater harm endorsement → greater moral self-image) or incongruence hypothesis (i.e., that the

mediational effect described by the moral motivation hypothesis is moderated by multiculturalism attitude congruence).

Moderated Mediation: Moral Conviction x Attitude Congruence → Non-Moral Justifications Endorsement → Harm Endorsement

As a reminder, the *moral primacy and incongruence hypotheses* state that stronger moral conviction should predict greater harm endorsement (both relatively mild and severe) against oppositional (but not likeminded) targets through non-moral justifications as mediators (i.e., advantageous comparison, euphemistic labeling, denial of consequences, and dehumanization; see Figure 4). That is, the direct effect of moral conviction on harm endorsement should be more strongly positive for opponents of multiculturalism who are attitudinally incongruent with the target (i.e., greater moral conviction → greater harm endorsement). In contrast, the direct effect should be non-significant or weaker for supporters of multiculturalism who are attitudinally congruent with the target. Furthermore, the indirect effect of moral conviction on harm endorsement through non-moral justification endorsement should be significant or stronger for those who are attitudinally incongruent with the target (i.e., greater moral conviction → greater non-moral justification endorsement → greater harm endorsement), but non-significant or weaker among those who are attitudinally congruent with the target.

Finally, if moral conviction is the primary driver of harm endorsement, then competing moderated mediation models placing each non-moral justification as the focal predictor rather than moral conviction should be non-significant (or have a worse model fit). These hypotheses were unsupported.

Once again, I used Hayes' (2018) PROCESS macro Model Template 8 with 5,000 bootstrapped resamplings to estimate two models, one for each kind of harm. I entered moral

conviction as the focal predictor; endorsement of advantageous comparison, euphemistic

labeling, denial of consequences, and dehumanization as mediators; endorsement of protest

behaviors (relatively mild/severe) as the dependent variables; and participants' multiculturalism

attitudes as a moderator of the A paths and the direct path (0 = attitudinal congruence with

target/supporters of multiculturalism; 1 = attitudinal incongruence with target/opponents of

multiculturalism).

Direct effects.

Mild harm endorsement. Contrary to the moral primacy and incongruence hypotheses,

the conditional direct effect of moral conviction on mild harm endorsement was non-significant

(see Table 24, especially the moral conviction x attitude congruence estimate).

Severe harm endorsement. Contrary to predictions, the *conditional* direct effect of moral

conviction on severe harm endorsement was non-significant. However, similar to Study 1, there

was a significant, unmoderated direct effect of moral conviction on severe harm endorsement,

such that greater moral conviction directly predicted weaker endorsement of severe harms (see

Table 25 for all direct effect estimates).

Indirect effects.

Moral conviction → Non-moral justification endorsement → Mild harm endorsement.

I next examined the indirect effect of moral conviction on moral mild harm endorsement through

non-moral justifications, and whether those paths were moderated by multiculturalism attitude

congruence (see Table 24 and Figure 12 for all path estimates). Contrary to hypotheses, greater

moral conviction predicted weaker euphemistic labeling, which in turn predicted greater

endorsement of mild harms: an effect that was unmoderated by multiculturalism attitude (index

of <u>moderated</u> mediation = -0.07, SE_{boot} = 0.05, 95% CI: [-0.19, 0.03]).[17] All other indirect effects

of moral conviction on harm endorsement through non-moral justifications were non-significant,

as were all remaining indices of moderated mediation.[18]

Beyond looking for evidence of indirect effects that supported my hypotheses, it is

interesting to note that, regardless of multiculturalism attitude congruence, greater moral

conviction significantly predicted *weaker* endorsement of dehumanization. Moreover, moral

conviction and attitude congruence interacted to predict denial of consequences: Greater moral

conviction predicted weaker denial of consequences among people who were attitudinally

congruent with the speaker/supporters of multiculturalism, B = -0.22, SE = 0.06, $t(386)$ = -3.86,

$p < 0.001$, but moral conviction did not predicted denial of consequences among people who

were attitudinally incongruent with the speaker/opponents of multiculturalism, B = 0.01, SE =

0.09, $t(386)$ = 0.15, $p = 0.88$. Lastly, although moral conviction did not predict advantageous

comparison, greater endorsement of advantageous comparison significantly predicted greater

endorsement of mild harms.

Overall, there was no evidence that the link between moral conviction and mild harm

endorsement was mediated by non-moral justifications in the predicted pattern (i.e., greater

moral conviction → greater non-moral justifications → greater harm endorsement). Like Study

1, because the moral primacy hypothesis was not supported, I did not analyze competing

---

[17] I confirmed that the indirect effect of moral conviction on mild harm endorsement through euphemistic labeling was significant in a mediation model that did not explore moderation by multiculturalism attitude congruence (i.e., Model Template 4).
[18] Advantageous comparison: index = -0.04, SE_{boot} = 0.04, 95% CI: [-0.12, 0.02]; denial of consequences: index = 0.05, SE_{boot} = 0.04, 95% CI: [-0.01, 0.14]; and dehumanization: index = 0.03, SE_{boot} = 0.03, 95% CI: [-0.02, 0.11].

moderated mediation models that would move each non-moral justification to the focal predictor

position in turn.

Table 24. *Moral Conviction → Non-Moral Justification Endorsement → Mild Harm Endorsement: All Path Estimates*

Predictor		Advantageous Comp (*M1*)		Mild Harm Endorsement (*Y*)
Moral conviction (*X*)	$a_{1Comp}\rightarrow$	0.00 (0.07), [-0.14, 0.13]	$c'_1 \rightarrow$	-0.09 (0.12), [-0.32, 0.14]
Participant multiculturalism attitude (*W*) 0 = Support, 1 = Oppose	$a_{2Comp}\rightarrow$	0.91 (0.52), [-0.10, 1.92]	$c'_2 \rightarrow$	0.27 (0.83), [-1.36, 1.91]
Advantageous Comp (*M1)*			$b_{Comp} \rightarrow$	**0.21 (0.09), [0.04, 0.38]***
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3Comp}\rightarrow$	-0.20 (0.13), [-0.47, 0.06]	$c'_3 \rightarrow$	0.01 (0.22), [-0.42, 0.43]
		Euphemistic Labeling (*M2*)		
Moral conviction (*X*)	$a_{1Euph}\rightarrow$	**-0.13 (0.04), [-0.22, -0.05]****	$c'_1 \rightarrow$	-0.09 (0.12), [-0.32, 0.14]
Participant multiculturalism attitude (*W*) 0 = Support, 1 = Oppose	$a_{2\,Euph} \rightarrow$	-0.24 (0.31), [-0.86, 0.37]	$c'_2 \rightarrow$	0.27 (0.83), [-1.36, 1.91]
Euphemistic Labeling (*M2*)			$b_{Euph} \rightarrow$	**-0.69 (0.17), [-1.02, -0.37]*****
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\,Euph} \rightarrow$	0.10 (0.08), [-0.06, 0.26]	$c'_3 \rightarrow$	0.01 (0.22), [-0.42, 0.43]

Denial of Consequences (*M3*)

Moral conviction (*X*)	$a_{1Deny} \rightarrow$	**-0.22 (0.06), [-0.33, -0.11]*****	$c'_1 \rightarrow$	-0.09 (0.12), [-0.32, 0.14]
Participant multiculturalism attitude (*W*) 0 = Support, 1 = Oppose	$a_{2\ Deny} \rightarrow$	-0.78 (0.42), [-1.59, 0.04]	$c'_2 \rightarrow$	0.27 (0.83), [-1.36, 1.91]
Denial of Consequences (*M3*)			$b_{Deny} \rightarrow$	0.21 (0.12), [-0.03, 0.44]
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\ Deny} \rightarrow$	**0.23 (0.11), [0.02, 0.44]***	$c'_3 \rightarrow$	0.01 (0.22), [-0.42, 0.43]

Dehumanization (*M4*)

Moral conviction (*X*)	$a_{1Dehum} \rightarrow$	**-0.21 (0.07), [-0.35, -0.07]****	$c'_1 \rightarrow$	-0.09 (0.12), [-0.32, 0.14]
Participant multiculturalism attitude (*W*) 0 = Support, 1 = Oppose	$a_{2\ Dehum} \rightarrow$	-0.05 (0.52), [-1.06, 0.97]	$c'_2 \rightarrow$	0.27 (0.83), [-1.36, 1.91]
Dehumanization (*M4*)			$b_{Dehum} \rightarrow$	0.15 (0.09), [-0.03, 0.32]
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\ Dehum} \rightarrow$	0.20 (0.13), [-0.07, 0.46]	$c'_3 \rightarrow$	0.01 (0.22), [-0.42, 0.43]

Note. ****$p < 0.001$; ***$p < 0.01$; *$p < 0.05$. Path coefficients are unstandardized, with standard errors represented in parentheses and 95% confidence intervals in brackets.
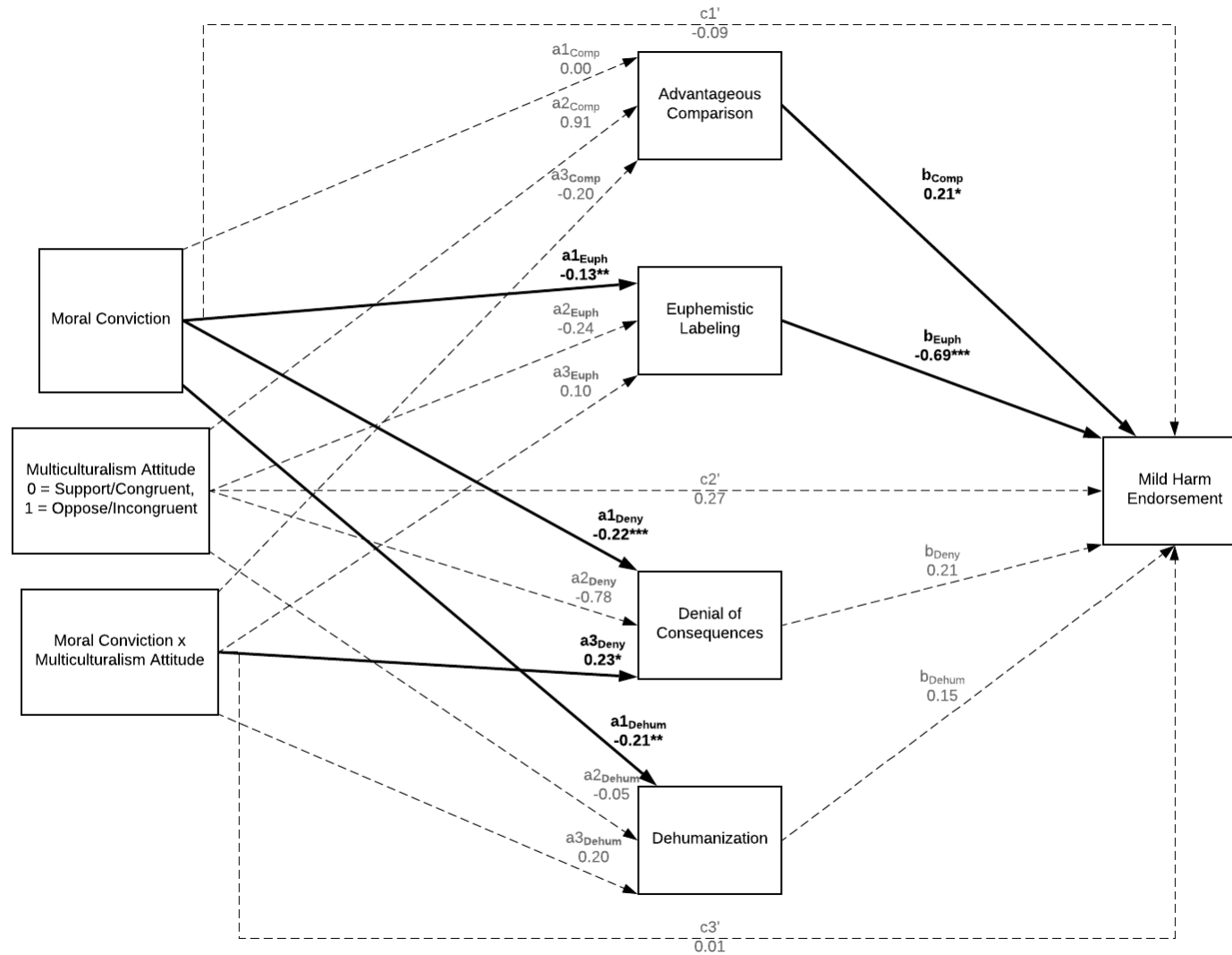
Figure 12. Statistical diagram of all path estimates represented in Table 24. Bolded paths are significant.

Moral conviction → Non-moral justification endorsement → Severe harm

endorsement. Next, I explored the indirect effects of moral conviction on severe harm

endorsement through non-moral justifications and whether multiculturalism attitude congruence

moderated these effects (see Table 25 and Figure 13 for all path estimates). Results revealed a

significant moderated mediation effect of moral conviction on severe harm endorsement through

denial of consequences (index of moderated mediation = 0.04, SE_{boot} = 0.02, 95% CI: [0.005,

0.10]). Contrary to hypotheses, the indirect effect was non-significant at the level of attitudinal

incongruence/opponents of multiculturalism, B = 0.00, SE_{boot} = 0.01, 95% CI: [-0.03, 0.03]. But

the indirect effect was significant among people who were attitudinally congruent with the

target/supporters of multiculturalism, B = -0.04, SE_{boot} = 0.02, 95% CI: [-0.08, -0.01] : Stronger

moral conviction predicted *weaker* denial of consequences, which in turn predicted weaker

endorsement of severe harms against likeminded targets.

Furthermore, contrary to the moral primacy hypothesis, results revealed that stronger

moral conviction significantly predicted weaker euphemistic labeling and dehumanization, which

in turn predicted weaker endorsement of severe harms. Both of these indirect effects were

unmoderated by multiculturalism attitude congruence, however, as indicated by null indices of

moderated mediation.[19][20]

Overall, because the moral primacy and incongruence hypotheses were not supported, I

did not explore competing moderated mediation models.

---

[19] Advantageous comparison: index = 0.00, SE_{boot} = 0.01, 95% CI: [-0.03, 0.02]; euphemistic
labeling: index = 0.03, SE_{boot} = 0.03, 95% CI: [-0.01, 0.10]; and dehumanization: index = 0.02,
SE_{boot} = 0.02, 95% CI: [-0.01, 0.06].
[20] Once again, I confirmed that the indirect effects of moral conviction on severe harm
endorsement through euphemistic labeling and dehumanization (unmoderated by
multiculturalism attitude congruence) were significant using PROCESS Macro Template 4.

Table 25. *Moral Conviction → Non-Moral Justification Endorsement → Severe Harm Endorsement: All Path Estimates*

Predictor		Advantageous Comp (*M1*)		Severe Harm Endorsement (*Y*)
Moral conviction (*X*)	$a_{1Comp}\rightarrow$	0.00 (0.07), [-0.14, 0.13]	$c'_1 \rightarrow$	**-0.11 (0.05), [-0.22, -0.01]***
Participant multiculturalism attitude (*W*) 0 = Support, 1 = Oppose	$a_{2Comp}\rightarrow$	0.91 (0.52), [-0.10, 1.92]	$c'_2 \rightarrow$	0.16 (0.38), [-0.59, 0.92]
Advantageous Comp (*M1*)			$b_{Comp} \rightarrow$	0.01 (0.04), [-0.07, 0.09]
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3Comp}\rightarrow$	-0.20 (0.13), [-0.47, 0.06]	$c'_3 \rightarrow$	0.04 (0.10), [-0.16, 0.23]
		Euphemistic Labeling (*M2*)		
Moral conviction (*X*)	$a_{1Euph}\rightarrow$	**-0.13 (0.04), [-0.22, -0.05]****	$c'_1 \rightarrow$	**-0.11 (0.05), [-0.22, -0.01]***
Participant multiculturalism attitude (*W*) 0 = Support, 1 = Oppose	$a_{2\ Euph} \rightarrow$	-0.24 (0.31), [-0.86, 0.37]	$c'_2 \rightarrow$	0.16 (0.38), [-0.59, 0.92]
Euphemistic Labeling (*M2*)			$b_{Euph} \rightarrow$	**0.35 (0.08), [0.20, 0.50]****
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\ Euph} \rightarrow$	0.10 (0.08), [-0.06, 0.26]	$c'_3 \rightarrow$	0.04 (0.10), [-0.16, 0.23]

Denial of Consequences (*M3*)

Moral conviction (*X*)	$a_{1\text{Deny}} \rightarrow$	**-0.22 (0.06), [-0.33, -0.11]*** **	$c'_1 \rightarrow$	**-0.11 (0.05), [-0.22, -0.01]***
Participant multiculturalism attitude (*W*) 0 = Support, 1 = Oppose	$a_{2\,\text{Deny}} \rightarrow$	-0.78 (0.42), [-1.59, 0.04]	$c'_2 \rightarrow$	0.16 (0.38), [-0.59, 0.92]
Denial of Consequences (*M3*)			$b_{\text{Deny}} \rightarrow$	**0.17 (0.05), [0.06, 0.28]*** **
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\,\text{Deny}} \rightarrow$	**0.23 (0.11), [0.02, 0.44]***	$c'_3 \rightarrow$	0.04 (0.10), [-0.16, 0.23]

Dehumanization (*M4*)

Moral conviction (*X*)	$a_{1\text{Dehum}} \rightarrow$	**-0.21 (0.07), [-0.35, -0.07]*** **	$c'_1 \rightarrow$	**-0.11 (0.05), [-0.22, -0.01]***
Participant multiculturalism attitude (*W*) 0 = Support, 1 = Oppose	$a_{2\,\text{Dehum}} \rightarrow$	-0.05 (0.52), [-1.06, 0.97]	$c'_2 \rightarrow$	0.16 (0.38), [-0.59, 0.92]
Dehumanization (*M4*)			$b_{\text{Dehum}} \rightarrow$	**0.09 (0.04), [0.01, 0.17]***
Moral conviction x Multiculturalism attitude (*X* x *W*)	$a_{3\,\text{Dehum}} \rightarrow$	0.20 (0.13), [-0.07, 0.46]	$c'_3 \rightarrow$	0.04 (0.10), [-0.16, 0.23]

Note. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. Path coefficients are unstandardized, with standard errors represented in parentheses and 95% confidence intervals in brackets.
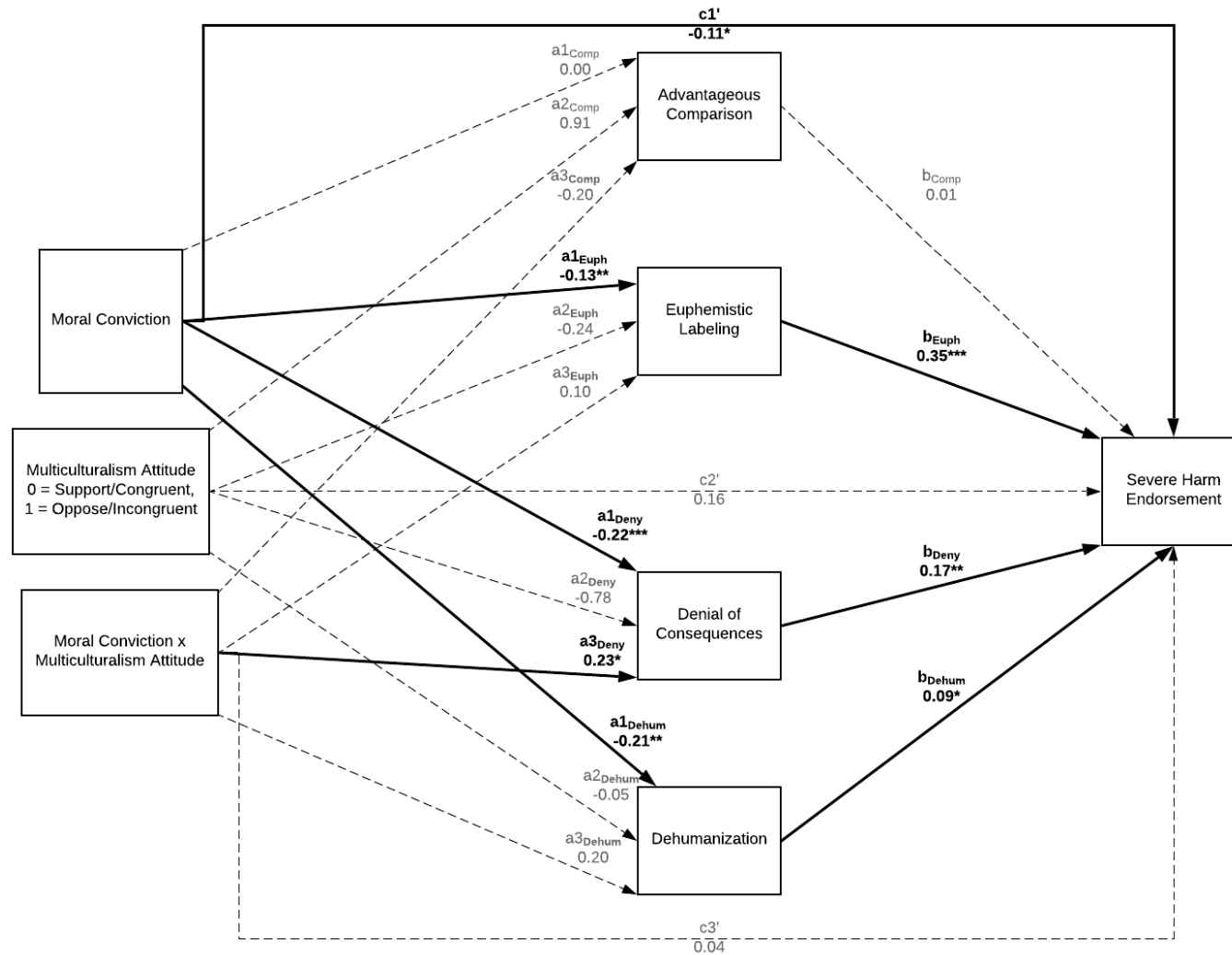
Figure 13. Statistical diagram of all path estimates represented in Table 25. Bolded paths are significant.

Summary. Overall, I found no support for the moral primacy or incongruence hypotheses, which state that greater moral conviction should predict greater harm endorsement through greater endorsement of non-moral justifications for oppositional (not likeminded) respondents. My results reveal that for <u>mild</u> protest behaviors, greater moral conviction predicted weaker euphemistic labeling, which in turn predicted greater endorsement of mild harms (an effect that was unmoderated by multiculturalism attitude congruence).[21] In contrast, for <u>severe</u> protest behaviors, greater moral conviction predicted weaker euphemistic labeling, denial of consequences[22], and dehumanization, which all in turn predicted weaker endorsement of severe harms (effects that were largely unmoderated by multiculturalism attitude congruence).[23] At a high level, these results are somewhat consistent with the findings from Study 1. These results suggest that, although people with strong moral conviction may be inclined to sometimes endorse mild harms at a multiculturalism protest, they are surprisingly *more* reluctant to severely harm others for that cause compared to people lower in moral conviction about multiculturalism.

Testing the Relationship Between Moral Conviction & Ethical Idealism

Once again, even though my hypotheses about the link between moral conviction and harm endorsement were unsupported, I nonetheless explored the relationship between moral conviction and ethical idealism. I predicted that stronger moral conviction about multiculturalism

---

[21] Contrasted against Study 1: In Study 1, I found a significant conditional direct effect of moral conviction on mild harm endorsement, such that greater moral conviction predicted greater mild harm endorsement for oppositional but not likeminded targets.

[22] Moral conviction predicted weaker denial of consequences among likeminded (not oppositional) respondents.

[23] Contrasted against Study 1: In Study 1, I found a significant direct effect of moral conviction on severe harm endorsement that was unmoderated by multiculturalism attitude congruence: Greater moral conviction predicted weaker endorsement of severe harms.

would be associated with weaker ethical idealism: the absolute belief that it is never acceptable to harm others in the pursuit of a noble cause.

Contrary to hypotheses, greater moral conviction about multiculturalism was associated with greater ethical idealism, $r(386) = 0.25$, $p < .001$. This positive effect of moral conviction on ethical idealism was marginally moderated by multiculturalism attitude congruence, $B = -0.24$, $SE = 0.13$, $t(384) = -1.91$, $p = 0.06$. Unpacking this marginal interaction revealed that, among supporters of multiculturalism who were attitudinally congruent with the target, greater moral conviction predicted greater ethical idealism, $B = 0.30$, $SE = 0.06$, $t(384) = 4.70$, $p < 0.001$. In contrast, moral conviction did not predict ethical idealism among opponents of multiculturalism who were attitudinally incongruent with the target, $B = 0.07$, $SE = 0.11$, $t(384) = 0.63$, $p = 0.53$. Overall, this result provides no support for my hypotheses.

Testing the Threat Hypothesis

For simplicity, I tested the threat hypothesis in the same way as Study 1: by dropping non-moral justifications as mediators and instead testing the three-way interaction of moral conviction, threat (continuously measured), and attitudinal incongruence with the target (0 = congruent; 1 = incongruent) predicting harm endorsement within the PROCESS macro using Model Template 3 and 5,000 bootstrapped resamplings. If the threat hypothesis is true, then I would expect to see that, among people who oppose multiculturalism/are attitudinally incongruent with the assigned target, stronger perception of threat should more strongly intensify the hypothesized positive association between moral conviction and endorsement of protest behaviors (both relatively mild and severe). In contrast, among participants who support multiculturalism/are attitudinally congruent with the assigned target, moral conviction should not

predict (or more weakly predict) endorsement of protest behaviors, regardless of perceived threat

at the rally.

Relatively mild protest behaviors. The overall model was not significant, $F(7, 382) =$

$1.34, p = 0.23, R^2 = 0.02$, nor was the predicted three-way interaction, B = 0.01, SE = 0.20,

$t(382) = 0.07, p = 0.95$. In fact, none of the terms in the model was significant, suggesting that

moral conviction, threat, and attitude congruence did not independently or multiplicatively

predict mild harm endorsement. In short, the threat hypothesis was unsupported for endorsement

of mild protest behaviors.

Relatively severe protest behaviors. The overall model was significant, $F(7, 382) =$

$5.64, p < 0.001, R^2 = 0.09$. However, none of the predictors in the model, nor their interaction

terms, were significant. Once again, the threat hypothesis was unsupported. [24]

Study 2 Discussion

Like Study 1, I did not find compelling support for my core moral motivation and moral

primacy hypotheses (see Table 26 for a summary of findings). Although the findings from Study

2 did not perfectly map onto what I found in Study 1, a couple similar themes emerged: (1)

There remains a link between moral conviction and mild harm endorsement, such that greater

moral conviction predicted greater mild harm endorsement (through weaker euphemistic labeling

in the case of Study 2); and (2) Greater moral conviction about multiculturalism once again

consistently predicted *weaker* endorsement of severe harms. At first blush, it appears that the

---

[24] Like Study 1, I also explored whether threat serves as a mediator in the link between moral
conviction and harmful behavior endorsement, using support/opposition to multiculturalism as a
moderator of this mediation model. Threat did not significantly mediate the link between moral
conviction and either kind of protest behavior (mild/severe), and these results will not be
explored further.

effect of moral conviction on harm endorsement is at least at a high level similar across both

studies.

Table 26. *Study 2 Summarized Findings*

Hypothesis	Multiculturalism Attitude Congruence	Expected	Observed	Hypothesis Supported?
Moral Motivation & Incongruence: **Mild** Harms	Incongruent Attitudes	Greater MC → Greater **mild** harms → Greater moral self-image	Non-significant links between MC, **mild** harms, moral self-image	No
	Congruent Attitudes	Weaker links between MC, **mild** harms, moral self-image than incongruent attitudes		
Moral Motivation & Incongruence: **Severe** Harms	Incongruent Attitudes	Greater MC → Greater **severe** harms → Greater moral self-image	Greater MC → Weaker **severe** harms → Weaker moral self-image (no moderation by attitude congruence)	No
	Congruent Attitudes	Weaker links between MC, **severe** harms, moral self-image than incongruent attitudes		
Moral Primacy & Incongruence: **Mild** Harms	Incongruent Attitudes	Greater MC → Greater non-moral justifications → Greater **mild** harms	Greater MC → Weaker **euphemistic labeling** → Greater **mild** harms (no moderation by attitude congruence)*	No
	Congruent Attitudes	Weaker links between MC, non-moral justifications, and **mild** harms than incongruent attitudes		
Moral Primacy & Incongruence: **Severe** Harms	Incongruent Attitudes	Greater MC → Greater non-moral justifications → Greater **severe** harms	Greater MC → Weaker non-moral justifications → Weaker **severe** harms* (inconsistent moderation by attitude congruence)	No
	Congruent Attitudes	Weaker links between MC, non-moral justifications, and **severe** harms than incongruent attitudes		

Threat: **Mild** Harms	Incongruent Attitudes	Threat x MC x Attitude Congruence interaction: Higher threat should intensify the MC → Harm link	No Threat x MC x Attitude Congruence Interaction	No
	Congruent Attitudes	No Threat x MC x Attitude Congruence interaction		
Threat: **Severe** Harms	Incongruent Attitudes	Threat x MC x Attitude Congruence interaction: Higher threat should intensify the MC → Harm link	No Threat x MC x Attitude Congruence Interaction	No
	Congruent Attitudes	No Threat x MC x Attitude Congruence interaction		

*_Note_. This is the primary result I derived from this particular analysis, although other paths of the model were significant.

Full Sample Analysis

Testing the Ideological Symmetry Hypothesis

The *ideological symmetry* hypothesis states that the relationship between moral conviction and harm toward political targets should be equivalent for respondents from both sides of the political aisle: both supporters and opponents of multiculturalism. If the *ideological symmetry hypothesis* is true, then I should have already observed that the results from Studies 1 and 2 are roughly equivalent. The ideological symmetry hypothesis argues that <u>all</u> people who experience attitudinal incongruence with a target should behave similarly as a function of their level of moral conviction about multiculturalism (i.e., supporters of multiculturalism in Study 1 & opponents multiculturalism in Study 2). Similarly, those who are congruent with the target should act equivalently (i.e., opponents of multiculturalism in Study 1 & supporters multiculturalism in Study 2). Overall, this reasoning suggests that the link between moral conviction and harm endorsement should be moderated by attitude congruence with the target, not side of issue per se.

However, I only observed conceptual similarities across these studies at a high level: overall directional similarities in the relationship between moral conviction and mild/severe harm endorsement. I did not observe strong similarities across studies in the mediational role of non-moral justifications, or consistent moderation by attitude congruence.

The next way I tested the *ideological symmetry hypothesis* was by building an aggregated, full-sample model that interacted moral conviction about multiculturalism, respondent attitude position on multiculturalism (ranging continuously from very much support to very much oppose), and attitude congruence with assigned target (congruent vs. incongruent)

to predict mild and severe harm endorsement.[25] If the ideological symmetry hypothesis is correct, then I should observe a *non-significant* three-way interaction between moral conviction, multiculturalism attitude position, and attitude congruence with the target. However, the two-way interaction between moral conviction and attitude congruence should be significant *regardless of respondent multiculturalism attitude position*. Independent of whether respondents support or oppose multiculturalism, greater moral conviction should more strongly predict greater harm endorsement among respondents who are <u>attitudinally incongruent</u> with their assigned target. In contrast, among respondents who are attitudinally congruent with their assigned target, greater moral conviction should not predict (or more weakly predict) harm endorsement.

To test the ideological symmetry hypothesis, I used PROCESS macro using Model Template 3 and 5,000 bootstrapped resamplings to estimate two models. For both models, I entered moral conviction as the focal predictor, multiculturalism attitude position as a moderator (continuously measured from -4 *very much support* to +4 *very much oppose*), and attitudinal incongruence with the assigned target as another moderator (0 = congruent; 1 = incongruent). The only way that the two models varied is in their dependent variable: the first model entered mild harm endorsement as the dependent variable, whereas the second entered severe harm endorsement.

Relatively mild protest behaviors. The overall model was significant, $F(7, 785) = 14.32$, $p < 0.001$, $R^2 = 0.11$. However, contrary to the ideological symmetry hypothesis, the

---

[25] Although I proposed to include non-moral justifications as mediators in this analysis, the inconsistent role that they played in Studies 1 and 2 led me to drop them for this analysis to simplify findings. Instead, I focus on understanding how multiculturalism attitude position moderates the primary variables in my experimental design (i.e., moral conviction, attitude congruence, harm endorsement).

three-way interaction between moral conviction about multiculturalism, respondent attitude position on multiculturalism, and attitude congruence with the assigned target was significant (see Table 27 for all model estimates). That is, the interactive effect of attitude congruence with the target and moral conviction on mild harm endorsement varied as a function of respondents' attitude position on multiculturalism.

More specifically, the simple interaction of moral conviction and attitude congruence was significant among those with a high level of support for multiculturism (-1 SD mean of multiculturalism attitude position). Among respondents who were assigned to evaluate a target who was attitudinally incongruent with them (i.e., for those at high levels of support for multiculturalism, this would correspond to obnoxious opponents of multiculturalism), stronger moral conviction predicted stronger endorsement of mild harms. In contrast, among those assigned to evaluate an attitudinally congruent target (i.e., obnoxious supporters of multiculturalism for those who strongly supported the issue), moral conviction did not significantly predict endorsement of mild protest behaviors.

Furthermore, the simple interaction of moral conviction and attitude congruence was significant among people with average attitude positions on multiculturalism, that is, people who moderately supported the issue. Greater moral conviction predicted greater endorsement of mild harms for people assigned to evaluate oppositional (but not likeminded) targets.

Finally, the simple interaction of moral conviction and attitude congruence was non-significant among people who opposed multiculturalism (+1 SD the mean of multiculturalism attitude position, which equated to slight opposition to the issue).

Taken together, the relationship between moral conviction, attitude congruence with the target, and mild harm endorsement varied as a function of respondents' attitude positions on

multiculturalism. Greater moral conviction predicted greater mild harm endorsement against oppositional (but not likeminded) targets, but only among people who strongly and averagely supported multiculturalism. Among people who slightly opposed multiculturalism, moral conviction did not predict mild harm endorsement. These results therefore contradict the ideological symmetry hypothesis.

Relatively severe protest behaviors. The overall model was significant, $F(7, 785) = 6.80$, $p < 0.001$, $R^2 = 0.06$. Supporting the ideological symmetry hypothesis, the three-way interaction between moral conviction, attitude congruence, and multiculturalism attitude position was non-significant, $B = 0.03$, $SE = 0.03$, $t(785) = 0.89$, $p = 0.37$. Interestingly, the two-way interaction between moral conviction and attitude congruence was marginally significant, $B = 0.16$, $SE = 0.09$, $t(785) = 1.75$, $p = 0.08$. Moreover, there was a main effect of moral conviction, such that greater conviction predicted *weaker* severe harm endorsement, $B = -0.24$, $SE = 0.07$, $t(785) = -3.57$, $p < 0.01$. No other effects were significant.

To probe the marginal two-way interaction between moral conviction and attitude congruence, I ran a new trimmed model using PROCESS Model Template 1 and 5,000 bootstrapped resamplings that included moral conviction about multiculturalism as the focal predictor, attitude congruence with the target as the moderator, severe harm endorsement as the dependent variable, and multiculturalism attitude position as a control variable (see Table 28 for all model estimates). Results revealed that, controlling for multiculturalism attitude position, greater moral conviction predicted *weaker* severe harm endorsement among people assigned to an attitudinally congruent (vs. incongruent) target.

Overall, these results support the ideological symmetry hypothesis, which states that the interactive effect of moral conviction and attitude congruence on harm endorsement should not

vary as a function of ideology. It is still important to emphasize, however, that these results run counter to the moral motivation hypothesis (i.e., greater moral conviction → greater harm endorsement for oppositional, not likeminded, targets). This full-sample analysis revealed that, independent of side of issue, greater moral conviction predicted weaker endorsement of severe harms against likeminded targets, and moral conviction did not predict harm endorsement for oppositional targets.

Table 27. *Moral Conviction x Attitude Congruence x Multiculturalism Attitude Position → Mild Harm Endorsement: All Path Estimates*

Effect	*F*	*p*	B	SE	*t*	*p*
MC x Att Congruence x Multiculturalism Position*			-0.14	0.05	*t*(785) = -2.55	0.01
MC x Att Congruence @ Multiculturalism High Support (- 1 SD Mean)*	*F*(1, 785) = 15.02	< 0.001				
MC x Att Congruence			-0.11	0.13	*t*(785) = -0.80	0.42
MC x Att Incongruence*			0.65	0.14	*t*(785) = 4.60	< 0.001
MC x Att Congruence @ Multiculturalism Mean Attitude Position	*F*(1, 785) = 12.80	< 0.01				
MC x Att Congruence			-0.09	0.10	*t*(785) = -0.92	0.36
MC x Att Inongruence*			0.41	0.10	*t*(785) = 4.14	< 0.001
MC x Att Congruence @ Multiculturalism Slight Oppose (+ 1 SD Mean)	*F*(1, 785) = 1.38	0.24				

*Note. *p* < 0.05; **p* < 0.01; ***p* < 0.001. Significant effects are bolded.*

Table 28. *Moral Conviction x Attitude Congruence → Severe Harm Endorsement (Controlling for Multiculturalism Attitudes): All Path Estimates*

Effect	B	SE	*t*	*p*
MC x Att Congruence^	0.13	0.08	*t*(788) = 1.70	0.09
MC x Att Congruence	-0.23	0.06	*t*(788) = -4.01	< 0.01
MC x Att Incongruence^	-0.10	0.06	*t*(788) = -1.78	0.08

Note. ^p < 0.10; **p* < 0.05; ***p* < 0.01; ****p* < 0.001. Significant effects are bolded.*

General Discussion

The overall goal of these studies was to understand the psychological levers that license people to harm others in contentious political contexts. Bandura (1999) proposed that a variety of factors provide justification for harm, including having a moral stake in the political issue at hand, comparing oneself to others who do even greater harms, labeling harms as less severe than they actually are, denying that the harms have lasting consequences, and dehumanizing the victims of harms. I theorized that *moral justification* would be the primary predictor of harm endorsement: If people could convince themselves that the harms served a higher moral purpose, they could transgress without jeopardizing their overall moral self-image (or they could even *boost* their moral self-images while transgressing, so long as they could reframe that behavior as a noble act). In contrast, non-moral justifications did not seem likely to provide a counteractive boost to people's moral self-image in the face of discrediting transgressions, and I therefore reasoned that they would be less likely to be primary drivers of harmful behavior compared to moral justification. Instead, I predicted that non-moral justifications would mediate the link between strong moral conviction for a cause and endorsement of harm against opponents of that cause.

Across two studies and over 800 participants, I found limited support for my theory of morally motivated harm endorsement. Next, I consider each of my major hypotheses one by one, and I identify which findings were supportive or oppositional to my predictions.

The *moral motivation* and *incongruence hypotheses* stated that stronger moral conviction would predict stronger harm endorsement for oppositional but not likeminded targets, which in turn would predict higher ratings of moral self-image. Providing partial support for the moral motivation and incongruence hypotheses, moral conviction *sometimes* licensed <u>mild</u> forms of

harm endorsement, such as picketing and staging sit-ins, even though this behavior did not in turn affect moral self-image ratings in the expected direction. Specifically, in Study 1, stronger moral conviction predicted greater endorsement of mild harms against oppositional but not likeminded protesters, which in turn predicted *weaker* (not stronger) moral self-image ratings. In contrast, in Study 2, moral conviction was not predictive of mild harm endorsement or moral self-image ratings regardless of attitude congruence with the target.

Further contradicting the moral motivation and incongruence hypotheses, across Studies 1 and 2, moral conviction consistently predicted *weaker* (not stronger) endorsement of severe harms (e.g., pushing, kicking, spitting at targets), regardless of participants' congruence with the target. In turn, this weaker support for severe harms went on to predict weaker moral self-image in Study 2 (but was unrelated to moral self-image in Study 1). Overall, these findings suggest that moral conviction is only sometimes positively related to harm endorsement[26], and harm endorsement is in turn never positively related to moral self-image ratings.

Next, the *moral primacy* and *incongruence hypotheses* stated that stronger moral conviction about multiculturalism would predict greater endorsement of non-moral justifications for oppositional (not likeminded) targets, which would in turn predict greater endorsement of harms. In both Studies 1 and 2, stronger moral conviction was linked to stronger endorsement of mild harms, even though there was no evidence that this link was consistently mediated by non-moral justifications in the predicted direction. Specifically, in Study 1, there was a direct positive effect of moral conviction on mild harm endorsement when the target was oppositional (not likeminded); this effect was unexpectedly not mediated by non-moral justifications. In contrast,

---

[26] When the harmful behavior was mild, and when respondents had a liberal position on the issue of multiculturalism; see the Ideological Symmetry hypothesis discussion below for more details.

in Study 2, stronger moral conviction predicted greater mild harm endorsement through *weaker* (not greater) euphemistic labeling, regardless of attitude congruence with the target. Overall, then, there was no consistent evidence that greater moral conviction predicted greater mild harm endorsement through greater endorsement of non-moral justifications.

Moreover, in both Studies 1 and 2, I found that stronger moral conviction consistently predicted *weaker* (not stronger) endorsement of <u>severe</u> harms regardless of attitude congruence with the target; in Study 1, this relationship between moral conviction and severe harm endorsement was direct, whereas in Study 2, this relationship was mediated by weaker endorsement of various non-moral justifications. Overall, these findings reveal that moral conviction was never positively linked to severe harm endorsement through non-moral justifications.

Interestingly, one commonality between mild and harm endorsement is that they were both predicted by euphemistic labeling. For mild harms, greater euphemistic labeling was associated with weaker harm endorsement. In contrast, for severe harms, greater euphemistic labeling was associated with greater severe harm endorsement. No other non-moral justification predicted *both* mild and severe harm endorsement across Studies 1 and 2.

Finally, the *ideological symmetry hypothesis* stated that the predicted relationships between moral conviction, attitude congruence with the target, and harm endorsement would be equivalent regardless of respondents' side of issue for multiculturalism. That is, I expected that, for people with both liberal and conservative positions on multiculturalism, moral conviction would positively predict harm endorsement for oppositional (not likeminded) targets. Results revealed that, although <u>severe</u> harm endorsement did not vary as a function of respondents' side of issue, judgments did vary for <u>mild</u> harms. Specifically, for supporters of multiculturalism with

liberal positions on the issue, greater moral conviction predicted greater mild harm endorsement for oppositional (but not likeminded) targets. In contrast, for opponents of multiculturalism with conservative positions on the issue, moral conviction and mild harm endorsement were unrelated regardless of attitude congruence with the target. Overall, it appears that people with more liberal positions on multiculturalism were more motivated by their strong moral convictions to endorse mild protest behaviors such as picketing and staging sit-ins.

In summary, the major conclusions from these studies are: (1) stronger moral conviction about multiculturalism consistently motivated stronger mild harm endorsement for oppositional but not likeminded targets, but only for people with liberal (not conservative) positions on the issue of multiculturalism; (2) stronger moral conviction about multiculturalism consistently *weakened* endorsement of severe harms, regardless of people's attitude congruence with the target or their attitudes about multiculturalism; and (3) non-moral justifications inconsistently mediated the relationships described above. In short, although people appear to be strategic in endorsing mild harms to serve their strong moral convictions, they are universally opposed to endorsing severe harms in the context of contentious political rallies—especially when they see the rally in moralized terms. Turning back toward Bandura's (1999) theory, these findings *do not* provide persuasive evidence that moral motivation is the primary route toward inflicting harm on opponents (at least in the context of Alt-Right rallies using the current methodology).

Relation to Neighboring Concepts

The Moral Mandate Effect. These results supplement research on the moral mandate effect: the tendency for people to be motivated to achieve a morally mandated outcome, even if the means of doing so are transgressive. Still unclear is how far people are willing to go to serve a noble cause. People are willing to break social norms for their strong moral convictions,

including norms related to conformity within social groups (Hornsey et al., 2003; Hornsey et al., 2007) and group cooperation (Skitka et al., 2005). Moral convictions also license transgressions that seem more extreme than those described above, including behavioral discrimination against opponents (Wright et al., 2008) and endorsement of vigilante justice for guilty defendants (Skitka & Houston, 2001). Very relevant to the current studies is the finding that strong moral convictions sometimes motivate people to support hostile forms of collective action against organizations that undermine those beliefs, such as defacing organizational property (Zaal et al., 2011).

However, there is an alternative stream of research that suggests that strong moral convictions do not *always* legitimize belief-bolstering transgressions; there appears to be a limit to the moral mandate effect. For example, in one set of studies, people were not supportive of others who committed heinous acts to serve a shared moral belief (e.g., bombing abortion clinics to bolster prolife beliefs; Mullen & Skitka, 2006).

The current studies help integrate the seemingly disparate findings described above by highlighting that transgression severity might be an important moderator of the moral mandate effect. Strong moral convictions may enable people to engage in harms that are perceived as mild: picketing, staging sit-ins, and possibly actions similar to those in Zaal et al.'s (2011) study, like punishing immoral organizations via hostile collective action. However, they do not appear to license harms that are perceived as more severe: pushing, kicking, spitting at opponents, or possibly those actions similar to Mullen & Skitka's (2006) study, like bombing an abortion clinic. Future research should explore the degree to which political behaviors are perceived as harmful before evaluating whether moral convictions predict people's willingness to endorse

them. If my reasoning is correct, harmfulness judgments of political behavior should moderate whether people are willing to enact them.

Virtuous violence. My theory of morally motivated violence is consistent with the concept of *virtuous violence*, even though my results are not. According to Rai and Fiske (2011), people sometimes perceive that inflicting harm on others is not only justifiable, but virtuous: specifically, when that violence helps sustain important social relationships. Using a cultural psychological approach, they argue that most people in most cultures perceive violence as morally right under some circumstances—even going so far as to applaud such acts. Common examples from their research included wartime violence, torture, and honor-related violence. However, given the argument that virtuous violence is observed nearly universally across cultures, an important question arises: Why did people resist endorsing severe acts of violence in the current studies, especially when they had a moral stake in the issue at hand?

One reason why I may not have observed morally motivated violence in the current studies is because the perceived social-relational context of the studies may not have supported such violence. Indeed, people's judgments of harm depend on the social-relational context that they find themselves in: communal sharing, authority ranking, equality matching, or market pricing (Rai & Fiske, 2011). Each of these social-relational contexts is associated with distinct motivations that are theorized to drive acceptance of violent behavior under unique circumstances. For example, when a communal sharing relational model is salient, people are motivated to protect the integrity of the in-group against threats (i.e., the unity motive). People within a communal sharing relational context are thought to accept violence when it preserves the integrity of a cohesive ingroup, for example, by aggressing against threatening outgroup members or tainted ingroup members who are believed to be contaminated. One could argue that

this relational context could be salient in contentious political rallies, to the extent that people

employ an "us versus them" mentality and perceive clear ingroup/outgroup boundaries. In

contrast, when an equality matching relational model is salient, people are motivated to enforce

an even balance in social relationships via reciprocity and eye-for-an-eye forms of revenge (i.e.,

the equality motive). People within an equality matching relational context are believed to accept

violence when it evens the score against others who committed equal harms against them.

Considering the relational models that were likely employed by participants in the current

studies could shed light on people's lack of acceptance of severe harm against opponents. It is

possible that people may not have engaged in a communal sharing model in my studies because

they may not have felt unified with other attitudinal ingroup members; this seems plausible given

how difficult it likely was for people to feel a sense of connection to imaginary ingroup members

at a rally that had not happened yet. Instead, people who participated in my studies may have

employed an equality matching relational model. An extension of the equality motive, then,

might have been an unwillingness to harm targets who did not directly harm oneself (i.e., eye-

for-an-eye style aggression, or lack thereof; see Rai and Fiske (2011) for other relational models

and motives).

Second, even though virtuous violence is observed across nearly all cultures, that does

not mean that it is easy to commit such acts. Rai and Fiske (2012) note that: "…committing

violence toward another human being can be difficult, requiring training, social support and

modeling, effort, practice, and experience before it becomes second nature…When people do

commit moral violence to others, they often feel guilt, shame, remorse, sadness, nausea, or horror

because of antiviolence motives that operate alongside the moral violence motives. But the fact

that people have competing motives to refrain from violence, yet often overcome those motives

to achieve virtuous violence, does not make their violence any less moral. It just means that humans are complex creatures, with many motives pushing in different directions" (p. 191). It is very likely that people who participated in my studies had many other competing motives that may have overwhelmed the morally motivated desire to inflict harm against opponents. For example, participants may have had a generalized motive to do no harm unto others (e.g., Crockett, Clark, Hauser, & Robbins, 2010), or a motive to respect civil liberties typically given to protesters at American rallies (Skitka, Liu, Yang, Chen, Liu, & Xu, 2012).[27] Future research should attempt to account for such competing motivations when trying to understand the link between moral conviction and violence endorsement. It is possible that individual differences in generalized harm aversion or respect for civil liberties could moderate the link between moral conviction and violence endorsement.

Limitations

This study has methodological limitations that may have contributed to people's comparative unwillingness to endorse severe (vs. mild) harms. First, because I used self-report measures to gauge people's endorsement of harm, social desirability bias could have obscured results. I attempted to reduce the likelihood that participants only shared social agreeable attitudes by including a variety of harm endorsement measures—some of which seemed more socially acceptable than personal support for violence (i.e., the degree that participants would encourage/discourage *other people* who enact the harmful behaviors; the degree that the harmful behaviors were justifiable/unjustifiable). Future research in this area should take care to employ

---

[27] Corroborating this idea, past work indicates that, although strong moral convictions are associated with political intolerance for Chinese respondents, strong moral commitments do not free Americans to endorse politically intolerant acts, such as restriction of free speech (Skitka et al., 2012).

methods that help curtail social desirability bias even further: for example, by administering scales designed to detect people who are socially biased responders or employing a bogus pipeline methodology (for a review, see Nederhof, 1985).

Second, my method involved administering an online survey outside of the context of a real rally; I asked participants to prospectively imagine how they would act and feel in a future context. Although my methodology offered the advantage of tightly controlling how the rally context was described to all participants, it was likely lacking in mundane realism. That is, the experimental session did not closely resemble what it would feel like to be on the ground at a contentious Alt-Right rally where violence could break out at any moment. Indeed, there are a number of important distinctions between real and hypothetical rallies that may account for a comparative lack of severe harm endorsement in the current studies: (1) little emotional engagement when considering hypothetical rallies vs. high emotionality at real rallies; (2) no opportunity to deindividuate when considering hypothetical rallies alone vs. the ability to deindividuate at real rallies within crowds; and (3) personal accountability when participating alone in a study about hypothetical rallies vs. diffusion of responsibility at real rallies among ingroup members. I attempted to maximize mundane realism by playing a video from the 2017 Unite the Rally for all participants before they considered a similar rally that was slated to occur in 2019. Nonetheless, it is important that future research supplement tightly-controlled laboratory studies with methodologies higher in external validity and mundane realism (see Future Directions for more details).

Third, because my method relied exclusively on self-report measures, participants may have been asked to report more than they could have had consciously known. This may have been especially problematic for my non-moral justification measures, to the extent that people

had little awareness of *why* they endorsed the harms that they did. Indeed, the idea that people are bad at reporting on their mental processes related to evaluation and judgment is not new in social psychology (e.g., Mandler, 1975; Miller, 1962; Neisser, 1967; cf. Nisbett & Wilson, 1977). More modern research suggests the same: that conscious introspection about why one made evaluations and judgments does not necessarily align with true underlying nonconscious mental processes. Rather than enabling people to dig deep enough to view unconscious mental processes, introspecting leads people to build explanations for their evaluations and judgments using consciously accessible self-knowledge. Importantly, these explanations may or may not be representative of one's true self/true processes; yet, people are unaware of this disconnect (for a review, see Wilson & Dunn, 2004). In the current studies, people may have had no conscious access to the extent to which each of Bandura's (1999) non-moral justifications shaped their harm judgments—or if other processes unexplored in the current study shaped those decisions. Future research employing a variety of methods beyond self-report measures may help reveal underlying reasons for harm endorsement (see Future Directions for more details).

Fourth, it is important to highlight one limitation of the measurement-of-mediation approach used in the current studies (Spencer, Zanna, & Fong, 2005). Specifically, because I tested mediation models using only measured (not manipulated) variables, other causal orders may fit the data better than the one implied by the current studies (i.e., moral conviction → non-moral justifications → harm endorsement). Future research should take care to explore other causal orders that are possible among these variables. For example, should future research find that non-moral justifications mediate the relationship between moral conviction and harm endorsement, it would be crucial to test alternative models as well. For example, it is also plausible that non-moral justifications may occur post-hoc, *after* the intent to harm decision.

Comparing the model fit of alternative causal orders against that which was theorized could shed light on which more appropriately describes observed relations between the variables in this research.

Finally, it is important to note that the two primary clusters of harmful protest behaviors emphasized in the current studies—mild and severe harms—may vary on other dimensions besides transgression severity. Other ways that these two kinds of harm could have varied include: (1) the number of people within the target group likely impacted by those harms (i.e., a few identifiable victims suffering a great deal as a result of behaviors like kicking and pepper spraying vs. a diffusion of harm across a whole target group for behaviors like picketing and staging sit-ins); (2) the degree to which the behaviors are perceived as normative vs. non-normative forms of collective action (Becker & Tausch, 2015); or (3) the degree to which the behaviors are perpetrated by just one aggressor vs. many. The implication of these alternative dimensions of harmful protest behaviors is that any of them could moderate people's endorsement of harms: *not* necessarily harm severity. Future research should attempt to ensure that mild vs. severe harms do not vary on other dimensions to strengthen claims that transgression severity moderates the link between moral conviction and people's willingness to harm others.

Future Directions

Re-examining the predictive power of Bandura's (1999) justifications for political violence using mixed methods. Future research should employ a variety of research methods to help probe predictors of harm at political rallies. First, field studies afford the opportunity to examine the processes proposed in my dissertation in real rather than hypothetical contexts, thus maximizing mundane realism and external validity. The current studies, for example, could be

replicated during a real political rally by interviewing protesters as they are immersed in a

political conflict. It is likely that administering my questionnaires during a political rally would

help enhance variance in harm endorsement; when people are genuinely riled up during a rally

and in true conflict with oppositional opponents, they may be more likely to experience morally

motivated harm motivations.

Second, qualitative methods could be employed to supplement survey methodologies for

the study of political violence. Namely, there is opportunity to gain a deeper understanding of

people's motivations for harm, and to evaluate whether those motivations align with Bandura's

(1999) theorized justifications. An interesting study idea, for example, could be to conduct semi-

structured qualitative interviews with people who are arrested for misconduct during real

political rallies. By potentially highlighting justifications beyond those examined in the current

studies, this kind of method could be useful for theory generation. Moreover, the relative

proportion of times that moral motivation, advantageous comparison, euphemistic labeling,

denial of consequences, and dehumanization are mentioned could also clarify which of

Bandura's (1999) processes are primary.

Third, archival methods could reveal historic motivations for political violence. For

example, to test the predictive power of moral conviction for political violence, future research

could code news story coverage of political rallies for moral content (e.g., by using the moral

foundations dictionary; Graham, Haidt, & Nosek, 2009). Then, researchers could correlate the

extent to which the rallies were framed in moral language with arrest records after the rallies

occurred. One might expect that the more that rallies are framed in moral language, the more

likely they would be to predict violent altercations between protesters and arrests. Or, perhaps

stronger moral language would predict greater crime severity during the rallies (if arrest records were coded as well for severity).

The role of previous activism experience in shaping people's reports & acceptance of morally motivated transgressions. It is possible that people's prior activism experience may moderate the link between moral conviction and harm endorsement, for a couple of reasons. First, compared to non-activists, people with a wealth of activism experience may have more accurate self-knowledge about their endorsement of harms and the processes driving those judgments, which could bode well for future field studies in this area utilizing self-report measures. Consistent with self-perception theory (Bem, 1972), activists (vs. non-activists) may be able to look to their own behavior (e.g., "How did I react to X rally behavior in the past?") to accurately project their endorsement of that behavior (e.g., "I must find X behavior acceptable to some extent if I did nothing to stop it from happening in the past"). In contrast, non-activists have no relevant behavioral data from which to anchor their judgments, and their self-knowledge about how they would feel about various harmful protest behaviors could be inaccurate as a result.

Moreover, it is possible that activists (vs. non-activists) may be comparatively desensitized to harmful protest behaviors and therefore more inclined to endorse them. The idea that people can become desensitized to violence over repeated exposures has been echoed prominently in the study of violent media (e.g., Carnagey, Anderson, & Bushman, 2006), and even in Bandura's (1999) theorizing of moral disengagement (termed "gradualistic moral disengagement"). People with no activism experience may be only comfortable with mild harms; but increasing attendance at contentious rallies may inflate their acceptance of increasingly severe harms. Given the relevance of my theory of moral motivation to political rallies, it is

important for future research to consider how harm endorsement varies as a function of activism experience (or whether past rally participation correlates positively in a linear fashion with harm endorsement, to suggest violence desensitization via activism).

Morally motivated transgressions from observers' perspectives. Finally, it may be possible to enhance variance in evaluations of harmful behavior by more deeply exploring *observers'* reactions to other people who commit such transgressions. Observers' attributions for a target's misdeeds are shaped by the extent that the action is consistent or inconsistent with a salient value (Morgan, Mullen, & Skitka, 2010), especially when those values are imbued with strong moral fervor (Mueller & Skitka, 2017). The more strongly observers agree with a normatively disengaged target's moral viewpoint, the more lenient their judgments are of that person (Mueller & Skitka, 2017; cf. Mullen & Skitka, 2006). Just like transgressors, morally likeminded observers may construe a target's actions as a courageous norm violation that serves a higher order good, that is, as a small price to pay for advancing a noble cause. Observers' own moral investment in the cause should therefore be an important factor that influences their attributions for targets committing morally motivated transgressions, even those that are harmful. Future research could explore the extent to which an observers' moral conviction predicts lenient judgments of likeminded (vs. oppositional) targets who do harmful acts at a rally to support a noble cause.

Conclusion

Beyond the world of Harry Potter, this area of research could have important implications: Over the last two decades, as ideological divides have grown in U.S. politics, so too has partisan antipathy (Pew, 2014). As just one example, conflict swelled at the 2017 Unite the Right Rally after white nationalists chanted intimidating Nazi and white supremacist slogans,

such as 'You will not replace us' and 'Jews will not replace us.' Counter-protesters chanted their own inflammatory slogans such as 'Kill All Nazis' and 'punch a Nazi in the mouth.' Typical behavior at the rally included pushing, punching, and spraying pepper spray into opponents' eyes. More lethal violence erupted when a self-identified white supremacist drove through a crowd of protesters, leaving one woman dead and several people injured.

Still largely unclear, however, are the psychological precursors that lead people to commit such acts of violence against political opponents. Yet, to reduce intergroup violence— like that observed at the Unite the Right Rally—we must gain a stronger understanding of the motives that drive it. How can people convince themselves that acts of violence are warranted? I attempted to answer this question in my dissertation studies. Although my studies did not illuminate why people endorse severe acts of violence, they did highlight the possible moderating role of transgression severity in people's willingness to harm. People seem to strategically endorse mild harms at political rallies, but they do not support severe harms even when they have a moral stake in the issue at hand. Future research employing a variety of methodologies should continue exploring the psychology of severe harm endorsement.

References

Aquino, K., & Reed, A. II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*, 1423–1440.

Aramovich, N. P., Lytle, B. L. & Skitka, L. J. (2012). Opposing torture: Moral conviction and resistance to majority influence, *Social Influence*, *1,* 21 - 34.

Bandura, A. (1989). Self-regulation of motivation and action through internal standards and goal systems. In L. A. Pervin (Ed.), *Goal concepts in personality and social psychology* (pp. 19-85). Hillsdale, NJ: Erlbaum.

Bandura, A. (1999). Moral disengagement in the perpetuation of inhumanities. *Personality and Social Psychology Review*, *3*(3), 193-209.

Bandura, A. (2011). Moral disengagement. In D. J. Christie (Ed.), *The encyclopedia of peace psychology*. Blackwell Publishing Ltd.

Bastian, B., & Haslam, N. (2010). Excluded from humanity: The dehumanizing effects of social ostracism. *Journal of Experimental Social Psychology*, *46*, 107-113.

Bastian, B., Denson, T. F., & Haslam, N. (2013). The roles of dehumanization and moral outrage in retributive justice. *PLOS One, 8*(4), 1-10.

Bauman, C. W., & Skitka, L. J. (2009). Moral conflict and procedural justice: Moral mandates as constraints to voice effects. *Australian Journal of Psychology, 61,* 40 – 49.

Becker, J. C., & Tausch, N. (2015). A dynamic model of engagement in normative and non-normative collective action: Psychological antecedents, consequences, and barriers. *European Review of Social Psychology*, *26*(1), 43-92.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, New York: Academic.

Blasi, A. (1983). Moral development and moral action: A theoretical perspective. *Developmental*

Review, 3, 178–210.

Blasi, A. (2004). Moral functioning: Moral understanding and personality. In D. K. Lapsley & D.

Narvaez (Eds.), *Moral development, self and identity: Essays in honor of Augusto Blasi*.

Mahwah, NJ: Lawrence Erlbaum & Associates.

Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The

ideological-conflict hypothesis: Intolerance among both liberals and conservatives.

Current Directions in Psychological Science, 23, 27-34.

Carnagey, N. L., Anderson, C. A., & Bushman, B. J. (2006). The effect of video game violence

on physiological desensitization to real-life violence. *Journal of Experimental Social

Psychology*, *43*(3), 489-496.

Chambers, J. R., Schlenker, B. R., & Collisson, B. (2013). Ideology and prejudice: The role of

value conflicts. *Psychological Science, 24*, 140-149.

Colby, A., & Damon, W. (1992). *Some do care: Contemporary lives of moral commitment.* New

York: Free Press.

Crawford, J. T. (2014). Ideological symmetries and asymmetries in political intolerance and

prejudice toward political activist groups. *Journal of Experimental Social Psychology,

55*, 284-298.

Crawford, J. T. (2017). Are conservatives more sensitive to threat than liberals? It depends on

how we define threat and conservatism. *Social Cognition*, *35*(4), 354-373.

Crawford, J. T., & Pilanski, J. M. (2014). Political intolerance, right and left. *Political

Psychology*, *35*(6), 841-851.

Crocker, J., & Wolfe, C. T. (2001). Contingencies of self-worth. *Psychological Review, 108*,

593–623.

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively

influences moral judgment and behavior through effects on harm aversion. *PNAS*,

107(40), 17433-17438.

Diener, E., Dineen, J., Endresen, K., Beaman, A. L., & Fraser, S. C. (1975). Effects of altered

responsibility, cognitive set, and modeling on physical aggression and deindividuation.

Journal of Personality and Social Psychology, *31*, 328-337.

Doris, J. M. (1998). Persons, situations, and virtue ethics. *Noûs, 32*, 504-530.

Edmonds, D. (2013). *Would you kill the fat man? The trolley problem and what your answer

tells us about right and wrong*. Princeton, NJ: Princeton University Press.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140.

Fiske, A. P., & Rai, T. S. (2014). *Virtuous violence: Hurting and killing to create, sustain, end,

and honor social relationships*. Cambridge University Press.

Forsyth, D. R. (1980). A taxonomy of ethical ideologies. *Journal of Personality and Social

Psychology*, *39*(1), 175-184.

Ginges, J., Atran, S., Medin, D., & Shikaki, K. (2007). Sacred bounds on rational resolution of

violent conflict. *Proceedings of the National Academy of Sciences, 111(18),* 7357 – 7360.

Goethals, G. R., & Darley, J. M. (1977). Social comparison theory: An attributional approach. In

J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical

perspectives* (pp. 259–278). Washington, DC: Hemisphere.

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*(5), 1029-46.

Greitemeyer, T., Fischer, P., Kastenmueller, A., & Frey, D. (2006). Civil courage and helping

behaviour: Differences and similarities. *European Psychologist, 11,* 90–98.

Greitemeyer, T., Osswald, S., Fischer, P., & Frey, D. (2007). Civil courage: Implicit theories, determinants, and measurement. *Journal of Positive Psychology*, *2*, 115–119.

Hart, D., Yates, M., Fegley, S., & Wilson, G. (1995). Moral commitment in inner-city adolescents. In M. Killen and D. Hart (Eds.), *Morality in everyday life: Developmental perspectives*. New York, NY: Cambridge University Press.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and condition process analysis: A regression-based approach* (2nd ed.). New York, NY: The Guilford Press.

Hornsey, M. J., Smith, J. R., & Begg, D. (2007). Effects of norms among those with moral conviction: Counter-conformity emerges on intentions but not behaviors. *Social Influence*, *2*(4), 244-268.

Hornsey, M. J., Majkut, L., Terry, D. J., & McKimmie, B. M. (2003). On being loud and proud: Non-conformity and counter-conformity to group norms. *British Journal of Social Psychology*, *42*(3), 319-335.

Jordan, J., Leliveld, M. C., & Tenbrunsel, A. E. (2015). The moral self-image scale: Measuring and understanding the malleability of the moral self. *Frontiers in Psychology*, *15*, 1-16.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin, 129*, 339-375.

Kelman, H. C. (1987). The political psychology of the Israeli-Palestinian conflict: How can we overcome the barriers to a negotiated solution? *Political Psychology, 8,* 347.

Kohlberg, L. (1975). The cognitive-developmental approach to moral education. *The Phi Delta Kappan, 56,* 670 – 677.

Koonz, C. (2003). *The Nazi conscience*. Cambridge, MA: The Belknap Press of Harvard

University Press.

Kraut, R. E. (1973). Effects of social labeling on giving to charity. *Journal of Experimental Social Psychology, 9*, 551-562.

Kteily, N., Bruneau, E., Waytz, A., & Cotterill, S. (2015). The ascent of man: Theoretical and empirical evidence for blatant dehumanization. *Journal of Personality and Social Psychology*, *109*(5), 901-931.

Lapsley, D.K., & Lasky, B. (2001). Prototypic moral character. *Identity,* 1(4), 345-363.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, *93*, 234-249.

Mandler, G. (1975). Consciousness: Respectable, useful and probably necessary. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium.* Hillsdale, N.J.: Erlbaum.

Marques, J. M., Yzerbyt, V. Y., & Leyens, J. P. (1988). The 'Black Sheep Effect': Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology*, *18*, 1-16.

Milgram, S. (1974). *Obedience to authority.* New York: Harper & Row.

Miller, G. A. (1962). *Psychology: The science of mental life.* New York: Harper & Row.

Miller, W. I. (2000). *The mystery of courage.* Cambridge, MS: Harvard University Press.

Miller, D. T., & Effron, D. A. (2010). Psychological license: When it is needed and how it functions. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 43, pp. 117-158). San Diego, CA: Academic Press/Elsevier.

Monin, B., & Jordan, A. H. (2009). The dynamic moral self: A social psychological perspective. In D. Narvaez & D. Lapsley (Eds.), *Personality, identity, and character: Explorations in*

moral psychology (pp. 341–354). New York: Cambridge University Press.

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology, 81*, 33-43.

Morgan, G. S., Mullen, E., & Skitka, L. J. (2010). When values and attributions collide: Liberals' and conservatives' values motivate attributions for alleged misdeeds. *Personality and Social Psychology Bulletin, 36*, 1241 – 1254.

Morgan, G. S., Skitka, L. J., & Lytle, B. (under review). Universally and objectively true: Psychological foundations of moral conviction.

Mueller, A. B., & Skitka, L. J. (2017). Liars, damned liars, and zealots: The effect of moral mandates on transgressive advocacy acceptance. *Social Psychological and Personality Science, 9*(6), 711-718.

Mullen, E., & Skitka, L. J. (2006). Exploring the psychological underpinnings of the moral mandate effect: Motivated reasoning, identification, or affect? *Journal of Personality and Social Psychology, 90,* 629 - 643.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, *15*, 263-280.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton- Century-Crofts.

Niesta Kayser, D., Greitemeyer, T., Fischer, P., & Frey, D. (2010). Why mood affects help giving, but not moral courage: Comparing two types of prosocial behavior. *European Journal of Social Psychology*, *40*, 1136-1157.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231-259.

Opotow, S. (1990). Moral exclusion and injustice: An introduction. *Journal of Social Issues, 46,*

1-20.

Osswald, S., Greitemeyer, T., Fischer, P., & Frey, D. (2010). What is moral courage? Definition, explication, and classification of a complex construct. In C. L. S. Pury, & S. J. Lopez (Eds.), *The psychology of courage: Modern research on an ancient virtue; the psychology of courage: Modern research on an ancient virtue* (pp. 149-164). American Psychological Association, Washington, DC.

Pew Research Center (2014). Political polarization in the American public. Retrieved from http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/

Proulx, T., & Heine, S. J. (2010). The frog in Kierkegaard's beer: Finding meaning in the threat-compensation literature. *Social and Personality Psychology Compass, 4*, 889-905.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*(1), 57-75.

Rai, T. S., & Fiske, A. P. (2012). Beyond harm, intention, and dyads: Relationship regulation, virtuous violence, and metarelational morality. *Psychological Inquiry*, *23*, 189-193.

Rai, T. S., Valdesolo, P., & Graham, J. (2017). Dehumanization increases instrumental violence, but not moral violence. *Proceedings of the National Academy of Sciences*, *114*(32), 8511-8516.

Reed, A., II, & Aquino, K. (2003). Moral identity and the circle of moral regard toward out-groups. *Journal of Personality and Social Psychology, 84*, 1270-1286.

Reed, A. II, Aquino, K., & Levy, E. (2007). Moral identity and judgments of charitable behaviors. *Journal of Marketing, 71*, 178-193.

Reifen Tagar, M., Morgan, G. S., Skitka, L., & Halperin, E. (2013). Moral conviction in the context of protracted intergroup conflict: When ideology matters in the Israeli-Palestinian

conflict. *European Journal of Social Psychology*.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36.

Royakkers, L., & Van Est, R. (2010). The cubicle warrior: The marionette of digitalized warfare. *Ethics and Information Technology, 12*(3), 289-296.

Ryan, T. J. (2017). No compromise: Political consequences of moralized attitudes. *American Journal of Political Science*, *61*(2), 409-423.

Shao, D., Aquino, K., & Freeman, D. (2008). Beyond moral reasoning: A review of moral identity research and its implications for business ethics. *Business Ethics Quarterly, 18*, 513-540.

Sibley, C. G., & Duckitt, J. (2008). Personality and prejudice: A meta-analysis and theoretical review. *Personality and Social Psychology Review, 12*, 248-279.

Skitka, L. J. (2002). Do the means always justify the ends or do the ends sometimes justify the means? A value protection model of justice reasoning. *Personality and Social Psychology Bulletin, 28,* 588 - 597.

Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass, 4,* 267 - 281.

Skitka, L. J. (2012). Moral convictions and moral courage: Common denominators of good and evil. In M. Mikulincer & P. Shaver (Eds.), *Social psychology of morality: Exploring the causes of good and evil (pp.* 349-365). Washington DC: American Psychological Association.

Skitka, L. J., & Houston, D. (2001). When due process is of no consequence: Moral mandates

and presumed defendant guilt or innocence. *Social Justice Research*, *14*, 305-326.

[Reprinted in T. Tyler (Ed.) (2005), *The International Library of Essays in Law and Society: Procedural Justice II* (pp. 525 – 533), Surrey, UK: Ashgate].

Skitka, L. J., & Mullen, E. (2002). Understanding judgments of fairness in a real-world political context: A test of the value protection model of justice reasoning. *Personality and Social Psychology Bulletin, 28,* 1419-1429.

Skitka, L. J. & Wisneski, D. C. (2011). Moral conviction and emotion. *Emotion Review, 3,* 328 – 330.

Skitka, L. J., Bauman, C. W., & Lytle, B. L. (2009). The limits of legitimacy: Moral and religious convictions as constraints on deference to authority. *Journal of Personality and Social Psychology, 97,* 567 – 578.

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, *88*(6), 895-917.

Skitka, L. J., Morgan, G. S., & Wisneski, D. C. (2015). Political orientation and moral conviction: A conservative advantage or an equal opportunity motivator of political engagement? In J. Forgas, W. Crano, & K. Fiedler (Eds.) *Social psychology and politics.* New York: Psychology Press.

Skitka, L. J., Liu, J., Yang, Y., Chen, H., Liu, L. & Xu, L. (2012). Exploring the cross-cultural generalizability and scope of morally motivated intolerance. *Social Psychological and Personality and Science*.

Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review, 99,* 3–21.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why

experiments are often more effective than mediational analyses in examining

psychological processes. *Journal of Personality and Social Psychology*, *89*(6), 845-851.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In

L. Berkowitz (Ed.), *Advances in experimental social psychology, Volume 21: Social

psychological studies of the self: Perspectives and programs* (pp. 261-302). San Diego,

CA: Academic Press.

Steele, C. M. (1999). The psychology of self-affirmation: Sustaining the integrity of the self. In

R. F. Baumeister (Ed.), *The self in social psychology: Key readings in social psychology*

(pp. 372-390). Philadelphia, PA: Psychology Press.

Struch, N., & Schwartz, S. H. (1989). Intergroup aggression: Its predictors and distinctness from

in-group bias. *Journal of Personality and Social Psychology, 56,* 364 –373.

Tetlock, P.E., Kristel, O.V., Elson, S.B., Green, M.C., & Lerner, J.S. (2000). The psychology of

the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals.

Journal of Personality and Social Psychology, 78(5): 853-870.

Tilker, H. A. (1970). Socially responsible behavior as a function of observer responsibility and

victim feedback. *Journal of Personality and Social Psychology*, *14*, 95-100.

Turiel, E. (1983). *The development of social knowledge: Morality and convention.* Cambridge,

England: Cambridge University Press.

Turiel, E., Hildebrandt, C., & Wainryb, C. (1991). Judging social issues: Difficulties,

inconsistencies, and consistencies. *Monographs of the Society for Research in Child

Development*, *56*(2), 1-116.

van Prooijen, J. W., Krouwel, A. P. M., Boiten, M., & Eendebak, L. (2015). Fear among the

extremes: How political ideology predicts negative emotions and out-group derogation. *Personality and Social Psychology Bulletin, 41*, 485-497.

Walker, L. J., & Hennig, K. H. (2004). Differing conceptions of moral exemplarity: Just, brave, and caring. *Journal of Personality and Social Psychology, 86*, 629–647.

Walker, L. J., & Frimer, J. A. (2007). Moral personality of brave and caring exemplars. *Journal of Personality and Social Psychology, 93,* 845–860.

Wetherell, G. A., Brandt, M. J., & Reyna, C. (2013). Discrimination across the ideological divide: The role of perceptions of value violations and abstract values in discrimination by liberals and conservatives. *Social Psychological and Personality Science*, *4*, 658–667.

Wheeler, L., Martin, R., & Suls, J. (1997). The proxy model of social comparison for self-assessment of ability. *Personality and Social Psychology Review, 1,* 54–61.

Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, *90*(2), 245-271.

Wilson, T. D. & Dunn, E. W. (2004). Self-knowledge: Its limits, value, and potential for improvement. *Annual Review of Psychology*, *55*, 493-518.

Wright, J. C., Cullum, J., & Schwab, N. (2008). The cognitive and affective dimensions of moral conviction: Implications for attitudinal and behavioral measures of interpersonal tolerance. *Personality and Social Psychology Bulletin*, *34*, 1461–1476.

Zaal, M. P., Van Laar, C., Stahl, T., Ellemers, N., & Derks, B. (2011). By any means necessary: The effects of regulatory focus and moral conviction on hostile and benevolent forms of collective action. *British Journal of Social Psychology*, *50*, 670–689.

Appendix A
Corpus of Protestor Behaviors

Verbal/Symbolic Harms

1. Chanting in unison to intimidate the target
2. Hurling insults at the target from a distance
3. Hurling insults at the target from close proximity
4. Screaming at the target from a distance
5. Screaming at the target from close proximity
6. Swearing at the target from a distance
7. Swearing at the target from close proximity
8. Giving the target the middle finger
9. Using offensive symbols or messages on clothing or signs

Physical Harms

1. Pushing the target lightly
2. Pushing the target hard
3. Throwing small objects at the target
4. Throwing large objects at the target
5. Using pepper spray to temporarily blind the target
6. Punching the target
7. Slapping the target
8. Kicking the target
9. Spitting at the target

Harm to Property

1. Vandalizing the target's property
2. Setting the target's property on fire
3. Throwing up barricades on the target's turf, so that they cannot congregate
4. Setting off a smoke or stink bomb on the target's turf

Non-Normative Passivity

1. Making a human chain to block the target from advancing
2. Staging a sit-in in a public space to send a message to the target
3. Engaging in a hunger strike to send a message to the target
4. Holding picket signs to show opposition to the target

Appendix B
Pilot Study Measures

Captcha

What number is above the red car? _____

Attitudes Toward Multiculturalism

Multiculturalism means that all different groups within society are treated equally and have the same rights, and their cultural perspectives are given equal value and status (e.g., Muslims/Christians, Blacks/Whites, gay and straight people, etc.).

Do you support or oppose multiculturalism in the U.S.?
- ☐ Support
- ☐ Neutral/ neither
- ☐ Oppose

If you had to say which way you lean, would you say you support or oppose multiculturalism in the U.S.? [for people who answer 'neutral/ neither' to first attitude question]
- ☐ Lean toward support
- ☐ Lean toward oppose
- ☐ Neutral/ neither

To what extent do you support multiculturalism in the U.S.? [for people who answer 'support' to first attitude question]
- ☐ Slightly
- ☐ Moderately
- ☐ Much
- ☐ Very much

To what extent do you oppose multiculturalism in the U.S.? [for people who answer 'oppose' to first attitude question]
- ☐ Slightly
- ☐ Moderately
- ☐ Much
- ☐ Very much

Political Orientation

Are your political beliefs generally liberal or conservative?
- ☐ Liberal
- ☐ Neutral/ neither
- ☐ Conservative

If you had to say which way you lean, would you say you are more liberal or conservative? [for people who answer 'neutral/ neither' to first political orientation question]
- ☐ Lean toward liberal
- ☐ Lean toward conservative
- ☐ Neutral/ neither

To what extent are your political beliefs liberal? [for people who answer 'liberal' to first political orientation question]
- ☐ Slightly
- ☐ Moderately
- ☐ Much
- ☐ Very much

To what extent are your political beliefs conservative? [for people who answer 'conservative' to first political orientation question]
- ☐ Slightly
- ☐ Moderately
- ☐ Much
- ☐ Very much

Alt-Right Rally Description

In 2019, political activists are gearing up for an Alt-Right rally that may be one of the most heated political rallies of our time. On one side are members of the Alt-Right, who rally to protect the rights of White people in the United States by opposing multiculturalism, immigration, and political correctness. Members of the other side, including the left-wing anti-fascist group known as Antifa, rally to protect the rights of minorities in the United States by supporting multiculturalism, lenient immigration laws, and political correctness rules. The 2019 rally is currently being planned for Charlottesville, Virginia. Largely considered to be the continuation of the Charlottesville 'Unite the Right' rally, the rally is perceived by some as a

tactic to normalize white nationalism. Members of Antifa—the left-wing anti-fascist group known to use force to try to silence white supremacists—are expected to gear up for what may be a reincarnation of the 2017 'Unite the Right' rally.

Alt-Right Rally Conflict Video

Rallies with both protestors and counter-protestors lead to conflict between members of opposing sides. Please watch the video below to see how conflict swelled during the 2017 Unite the Right Rally in Charlottesville, Virginia.

[https://www.youtube.com/watch?v=26lkrCzObpQ&feature=youtu.be]

What are your impressions of the video you just watched?

To what extent do you think that the Unite the Right rally depicted in the video was violent?
☐ Not at all
☐ Slightly
☐ Moderately
☐ Much
☐ Very much

To what extent do you think that the protesters' behavior in the video was obnoxious?
☐ Not at all
☐ Slightly
☐ Moderately
☐ Much
☐ Very much

Harm Endorsement Block

Imagine you attend the upcoming Alt-Right rally and counter-protest. What kind of actions would you be comfortable taking against [pipe in opposing stance: supporters/opponents of multiculturalism] at the rally who display obnoxious behavior like that depicted in the video? That is, to what degree do you feel this behavior is an appropriate reaction against [supporters/opponents of multiculturalism] who display obnoxious behavior like that depicted in the video? For the remainder of today's study, you will be asked to consider your reactions to behaviors between protesters.

To what extent can you imagine a set of circumstances where an appropriate reaction for protesters like you would be to enact the following behaviors toward [pipe in opposing ideology: supporters/opponents of multiculturalism] who display obnoxious behavior like that depicted in the video?

	Not at all	Slightly	Moderately	Much	Very much
Chanting in unison to intimidate obnoxious [supporters/opponents of multiculturalism]					
Hurling insults at obnoxious [supporters/opponents of multiculturalism] from a distance					
Hurling insults at obnoxious [supporters/opponents of multiculturalism] from close proximity					
Screaming at obnoxious [supporters/opponents of multiculturalism] from a distance					
Screaming at obnoxious [supporters/opponents of multiculturalism] from close proximity					
Swearing at obnoxious [supporters/opponents of multiculturalism] from a distance					
Swearing at obnoxious [supporters/opponents of multiculturalism] from close proximity					
Giving obnoxious [supporters/opponents of multiculturalism] the middle finger					
Using offensive symbols or messages on clothing or signs aimed at obnoxious [supporters/opponents of multiculturalism]					
Pushing obnoxious [supporters/opponents of multiculturalism] lightly					
Pushing obnoxious [supporters/opponents of multiculturalism] hard					
Throwing small objects at obnoxious [supporters/opponents of multiculturalism]					

Throwing large objects at obnoxious [supporters/opponents of multiculturalism]				
Using pepper spray to temporarily blind obnoxious [supporters/opponents of multiculturalism]				
Punching obnoxious [supporters/opponents of multiculturalism]				
Slapping obnoxious [supporters/opponents of multiculturalism]				
Kicking obnoxious [supporters/opponents of multiculturalism]				
Spitting at obnoxious [supporters/opponents of multiculturalism]				
Vandalizing the property of obnoxious [supporters/opponents of multiculturalism]				
Setting the property of obnoxious [supporters/opponents of multiculturalism] on fire				
Throwing up barricades to block obnoxious [supporters/opponents of multiculturalism] from congregating				
Setting off a smoke or stink bomb near obnoxious [supporters/opponents of multiculturalism]				
Making a human chain to block obnoxious [supporters/opponents of multiculturalism] from advancing				
Staging a sit-in in a public space to send a message to obnoxious [supporters/opponents of multiculturalism]				
Engaging in a hunger strike to send a message to obnoxious [supporters/opponents of multiculturalism]				
Holding picket signs to show opposition to obnoxious				

| [supporters/opponents of multiculturalism] | | | | | |

Justifications & Threat Block

Thinking about your responses overall, how much do you agree that each factor below influenced your perceptions of these behaviors?

To what extent do you think…					
	Not at all	**Slightly**	**Moderately**	**Much**	**Very much**
Other protesters would normally behave worse than this					
Other people will do worse things than this at the upcoming rally					
These behaviors are better than how protesters normally act at rallies					
No one else I know would behave in these ways					
These behaviors are worse than how people normally act at rallies					
Given that other people will probably do worse things at the rally, these behaviors are not that bad					
As long as other people do worse things at the rally, these behaviors are acceptable					
These behaviors are just part of the fun at rallies					
These behaviors are more of a joke than serious					
These behaviors merely reflect healthy competition between people attending rallies					
It is important to take these behaviors seriously					
These behaviors are no laughing matter					
People who take issue with these kinds of behaviors can't take a joke					
These behaviors make rallies exciting, like a sport					
No one would actually be harmed by these behaviors					

These behaviors won't have lasting consequences					
No one would be seriously affected by these behaviors					
These behaviors can be harmful					
The consequences of behaviors like these can be severe					
The harms associated with these behaviors can negatively affect a person for a long time					
These behaviors are not physically or psychologically harmful					
Obnoxious [supporters/opponents of multiculturalism] are barbaric					
Obnoxious [supporters/opponents of multiculturalism] lack self-restraint, like animals					
Obnoxious [supporters/opponents of multiculturalism] are unsophisticated					
No one deserves to be treated like animals, not even obnoxious [supporters/opponents of multiculturalism]					
Obnoxious [supporters/opponents of multiculturalism] are refined and cultured					
Obnoxious [pipe in opposing stance: supporters/opponents of multiculturalism] are rational and logical, like they are intelligent					
Obnoxious [pipe in opposing stance: supporters/opponents of multiculturalism] are less than human, like animals					
The rally will violate your core political values and beliefs					
Your political values and beliefs will be undermined at the upcoming rally					
The rally is dangerous for society					
The upcoming rally will threaten democracy in the United States					
American values are at stake at the upcoming rally					
The upcoming rally is a threat to American culture					

The upcoming rally will make American society more dangerous					

Harmfulness Rating Block

To what extent are the following behaviors harmful to [pipe in opposing stance: supporters/opponents of multiculturalism] who display obnoxious behavior like that depicted in the video?					
	Not at all	Slightly	Moderately	Much	Very much
Chanting in unison to intimidate obnoxious [supporters/opponents of multiculturalism]?					
Hurling insults at obnoxious [supporters/opponents of multiculturalism] from a distance					
Hurling insults at obnoxious [supporters/opponents of multiculturalism] from close proximity					
Screaming at obnoxious [supporters/opponents of multiculturalism] from a distance					
Screaming at obnoxious [supporters/opponents of multiculturalism] from close proximity					
Swearing at obnoxious [supporters/opponents of multiculturalism] from a distance					
Swearing at obnoxious [supporters/opponents of multiculturalism] from close proximity					
Giving obnoxious [supporters/opponents of multiculturalism] the middle finger					
Using offensive symbols or messages on clothing or signs aimed at obnoxious [supporters/opponents of multiculturalism]					
Pushing obnoxious [supporters/opponents of multiculturalism] lightly					

Pushing obnoxious [supporters/opponents of multiculturalism] hard					
Throwing small objects at obnoxious [supporters/opponents of multiculturalism]					
Throwing large objects at obnoxious [supporters/opponents of multiculturalism]					
Using pepper spray to temporarily blind obnoxious [supporters/opponents of multiculturalism]					
Punching obnoxious [supporters/opponents of multiculturalism]					
Slapping obnoxious [supporters/opponents of multiculturalism]					
Kicking obnoxious [supporters/opponents of multiculturalism]					
Spitting at obnoxious [supporters/opponents of multiculturalism]					
Vandalizing the property of obnoxious [supporters/opponents of multiculturalism]					
Setting the property of obnoxious [supporters/opponents of multiculturalism] on fire					
Throwing up barricades to block obnoxious [supporters/opponents of multiculturalism] from congregating					
Setting off a smoke or stink bomb near obnoxious [supporters/opponents of multiculturalism]					
Making a human chain to block obnoxious [supporters/opponents of multiculturalism] from advancing					
Staging a sit-in in a public space to send a message to obnoxious [supporters/opponents of multiculturalism]					

Engaging in a hunger strike to send a message to obnoxious [supporters/opponents of multiculturalism]					
Holding picket signs to show opposition to obnoxious [supporters/opponents of multiculturalism]					

Impressions of Multiculturalism Block

Multiculturalism means that all different groups within society are treated equally and have the same rights, and their cultural perspectives are given equal value and status (e.g., Muslims/Christians, Blacks/Whites, gay and straight people, etc.).

To what extent do you perceive <u>multiculturalism</u> to be a key issue at stake at <u>Alt-Right/Antifa rallies</u>?

☐ Not at all
☐ Slightly
☐ Moderately
☐ Much
☐ Very much

To what extent do you perceive **support for multiculturalism** as a liberal or conservative position?

☐ Very liberal
☐ Moderately liberal
☐ Slightly liberal
☐ Neither liberal nor conservative
☐ Slightly conservative
☐ Moderately conservative
☐ Very conservative

To what extent do you perceive **opposition to multiculturalism** as a liberal or conservative position?

☐ Very liberal
☐ Moderately liberal
☐ Slightly liberal
☐ Neither liberal nor conservative
☐ Slightly conservative
☐ Moderately conservative
☐ Very conservative

Demographics

Please provide us with a bit of information about yourself.

1. Gender:
   - Male
   - Female

2. Age:

3. Is English your primary language?
   - Yes
   - No

4. Ethnicity (check all that apply)
   - ☐ African American
   - ☐ Asian or Pacific Islander
   - ☐ Caucasian (White)
   - ☐ Latino
   - ☐ Native American
   - ☐ Other

5. What is the highest level of education you have completed?
   - Less than high school
   - High school/GED
   - Some college
   - 2-year college degree
   - 4-year college degree
   - Master's degree
   - Doctoral degree

Attention Check Block

To what extent should we trust your data?
   - ☐ Not at all
   - ☐ Slightly
   - ☐ Moderately
   - ☐ Much
   - ☐ Very much

How distracted were you while completing this study?
   - ☐ Not at all
   - ☐ Slightly
   - ☐ Moderately
   - ☐ Much
   - ☐ Very much

To what extent were you paying attention in this study?

☐ Not at all
☐ Slightly
☐ Moderately
☐ Much
☐ Very much

What is the topic of this study? Sometimes participants do not carefully read the instructions. To correctly answer this question, please select the option "other" and write down the name of your favorite movie.

☐ Sexism
☐ Multiculturalism
☐ I don't remember
☐ Other (please specify:_____)

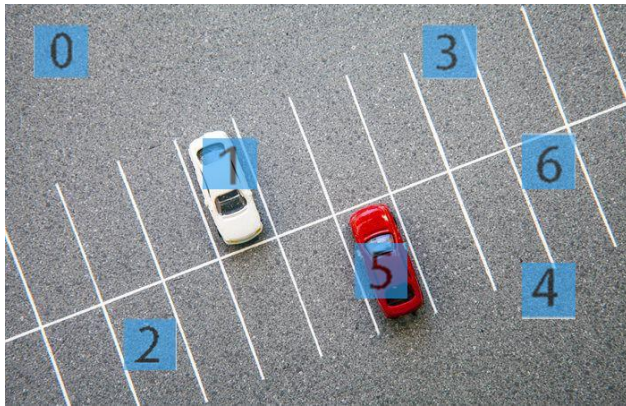Do you have any other thoughts about today's study?

Appendix C
Studies 1 & 2 Measures

Captcha

What number is above the red car? _____

Attitudes Toward Multiculturalism

Multiculturalism means that all different groups within society are treated equally and have the same rights, and their cultural perspectives are given equal value and status (e.g., Muslims/Christians, Blacks/Whites, gay and straight people, etc.).

Do you support or oppose multiculturalism in the U.S.?
☐ Support
☐ Neutral/ neither
☐ Oppose

If you had to say which way you lean, would you say you support or oppose multiculturalism in the U.S.? [for people who answer 'neutral/ neither' to first attitude question]
☐ Lean toward support
☐ Lean toward oppose

☐ Neutral/ neither

To what extent do you support multiculturalism in the U.S.? [for people who answer 'support' to first attitude question]
☐ Slightly
☐ Moderately
☐ Much
☐ Very much

To what extent do you oppose multiculturalism in the U.S.? [for people who answer 'oppose' to first attitude question]
☐ Slightly
☐ Moderately
☐ Much
☐ Very much

Attitude Strength, & Moral Conviction for Multiculturalism

Multiculturalism means that all different groups within society are treated equally and have the same rights, and their cultural perspectives are given equal value and status (e.g., Muslims/Christians, Blacks/Whites, gay and straight people, etc.).

To what extent is your position on <u>multiculturalism in the U.S.</u>…					
	Not at all	**Slightly**	**Moderately**	**Much**	**Very much**
something that you care a lot about?					
personally important to you?					
something you are certain about?					
something you are sure you are correct about?					
a reflection of your core moral beliefs and convictions?					
connected to your beliefs about fundamental right and wrong?					
based on moral principle?					

a moral stance?					

Political Orientation

Are your political beliefs generally liberal or conservative?
- ☐ Liberal
- ☐ Neutral/ Neither
- ☐ Conservative

If you had to say which way you lean, would you say you are more liberal or conservative? [for people who answer 'neutral/ neither' to first political orientation question]
- ☐ Lean toward liberal
- ☐ Lean toward conservative
- ☐ Neutral/ Neither

To what extent are your political beliefs liberal?
[for people who answer 'liberal' to first political orientation question]
- ☐ Slightly
- ☐ Moderately
- ☐ Much
- ☐ Very much

To what extent are your political beliefs conservative?
[for people who answer 'conservative' to first political orientation question]
- ☐ Slightly
- ☐ Moderately
- ☐ Much
- ☐ Very much

Alt-Right Rally Description

In 2019, political activists are gearing up for an Alt-Right rally that may be one of the most heated political rallies of our time. On one side are members of the Alt-Right, who rally to protect the rights of White people in the United States by opposing multiculturalism, immigration, and political correctness. Members of the other side, including the left-wing anti-fascist group known as Antifa, rally to protect the rights of minorities in the United States by supporting multiculturalism, lenient immigration laws, and political correctness rules. The 2019 rally is currently being planned for Charlottesville, Virginia. Largely considered to be the continuation of the Charlottesville 'Unite the Right' rally, the rally is perceived by some as a tactic to normalize white nationalism. Members of Antifa—the left-wing anti-fascist group known to use force to try to silence white supremacists—are expected to gear up for what may be a reincarnation of the 2017 'Unite the Right' rally.

Alt-Right Rally Conflict Video

Rallies with both protestors and counter-protestors lead to conflict between members of opposing sides. Please watch the video below to see how conflict swelled during the 2017 Unite the Right Rally in Charlottesville, Virginia.

[https://www.youtube.com/watch?v=26lkrCzObpQ&feature=youtu.be]

Harm Endorsement Block

Imagine you attend the upcoming Alt-Right/Antifa rally. What kind of actions would you be comfortable or uncomfortable with your side taking against [manipulated between-subjects: **opponents vs. supporters of multiculturalism**] who display obnoxious behavior like that depicted in the video? What actions would you **support or oppose** using against [**opponents vs. supporters of multiculturalism**] who attend the rally and display obnoxious behavior like that depicted in the video?

To what extent would you **support or oppose** using the following actions against obnoxious [**opponents vs. supporters of multiculturalism**] at the upcoming rally?							
	Strongly Support	**Moderately Support**	**Slightly Support**	**Neither Support nor Oppose**	**Slightly Oppose**	**Moderately Oppose**	**Strongly Oppose**
Pushing obnoxious [opponents vs. supporters of multiculturalism] lightly							
Pushing obnoxious [opponents vs. supporters of multiculturalism] hard							

Throwing small objects at obnoxious [opponents vs. supporters of multiculturalism]							
Using pepper spray to temporarily blind obnoxious [opponents vs. supporters of multiculturalism]							
Kicking obnoxious [opponents vs. supporters of multiculturalism]							
Spitting at obnoxious [opponents vs. supporters of multiculturalism]							
Staging a sit-in in a public space to send a message to obnoxious [opponents vs. supporters of multiculturalism]							
Holding picket signs to show opposition to obnoxious [opponents vs. supporters of multiculturalism]							

To what extent would you **encourage or discourage** others who use the following actions against obnoxious [**opponents vs. supporters of multiculturalism**] at the upcoming rally?							
	Strongly Encourage	**Moderately Encourage**	**Slightly Encourage**	**Neither Encourage nor Discourage**	**Slightly Discourage**	**Moderately Discourage**	**Strongly Discourage**
Pushing obnoxious [opponents vs. supporters of multiculturalism] lightly							
Pushing obnoxious [opponents vs. supporters of multiculturalism] hard							
Throwing small objects at obnoxious [opponents vs. supporters of multiculturalism]							
Using pepper spray to temporarily blind obnoxious [opponents vs. supporters of multiculturalism]							

Kicking obnoxious [opponents vs. supporters of multiculturalism]							
Spitting at obnoxious [opponents vs. supporters of multiculturalism]							
Staging a sit-in in a public space to send a message to obnoxious [opponents vs. supporters of multiculturalism]							
Holding picket signs to show opposition to obnoxious [opponents vs. supporters of multiculturalism]							

To what extent do you think the following actions are **justifiable or unjustifiable** to use against obnoxious [**opponents vs. supporters of multiculturalism**] at the upcoming rally?							
	Very Justifiable	**Moderately Justifiable**	**Slightly Justifiable**	**Neither Justifiable nor Unjustifiable**	**Slightly Unjustifiable**	**Moderately Unjustifiable**	**Very Unjustifiable**
Pushing obnoxious [opponents vs. supporters of multiculturalism] lightly							
Pushing obnoxious [opponents vs. supporters of multiculturalism] hard							
Throwing small objects at obnoxious [opponents vs. supporters of multiculturalism]							
Using pepper spray to temporarily blind obnoxious [opponents vs. supporters of multiculturalism]							
Kicking obnoxious [opponents vs. supporters of multiculturalism]							
Spitting at obnoxious [opponents vs. supporters of multiculturalism]							

Staging a sit-in in a public space to send a message to obnoxious [opponents vs. supporters of multiculturalism]						
Holding picket signs to show opposition to obnoxious [opponents vs. supporters of multiculturalism]						

Justifications & Threat Block

 Next, we are interested in learning more about how you decided which behaviors to support or oppose in the previous task. Thinking about your responses overall, how much do you agree that each factor below influenced your decision to support/oppose these behaviors?

To what extent do you think…					
	Not at all	**Slightly**	**Moderately**	**Much**	**Very much**
Other protesters would normally behave worse than this					
Other people will do worse things than this at the upcoming rally					
Other people will be more obnoxious than this					
Other people will be more harmful than this					
These behaviors are just part of the fun at rallies					
These behaviors are more of a joke than serious					
People who take issue with these kinds of behaviors can't take a joke					
These behaviors make rallies exciting, like a sport					

No one would actually be harmed by these behaviors					
These behaviors won't have lasting consequences					
No one would be seriously affected by these behaviors					
These behaviors are not physically or psychologically harmful					
Obnoxious [opponents vs. supporters of multiculturalism] are barbaric					
Obnoxious [opponents vs. supporters of multiculturalism] lack self-restraint, like animals					
Obnoxious [opponents vs. supporters of multiculturalism] are unsophisticated					
Obnoxious [opponents vs. supporters of multiculturalism] are less than human, like animals					
The rally will violate your core political values and beliefs					
Your political values and beliefs will be undermined at the upcoming rally					
The rally is dangerous for society					
The upcoming rally will threaten democracy in the United States					
American values are at stake at the upcoming rally					
The upcoming rally is a threat to American culture					
The upcoming rally will make American society more dangerous					

Moral Self-Image Block

Next, we're interested in learning about your personality and how you see yourself. Please respond to the following statements as they apply to you.

Compared to the <u>caring</u> person I want to be, I am…
- ☐ **Much less** caring than the person I want to be
- ☐ **Moderately less** caring than the person I want to be
- ☐ **Slightly less** caring than the person I want to be
- ☐ **Exactly as** caring as the person I want to be
- ☐ **Slightly more** caring than the person I want to be
- ☐ **Moderately more** caring than the person I want to be
- ☐ **Much more** caring than the person I want to be

Compared to the <u>compassionate</u> person I want to be, I am…
- ☐ **Much less** compassionate than the person I want to be
- ☐ **Moderately less** compassionate than the person I want to be
- ☐ **Slightly less** compassionate than the person I want to be
- ☐ **Exactly as** compassionate as the person I want to be
- ☐ **Slightly more** compassionate than the person I want to be
- ☐ **Moderately more** compassionate than the person I want to be
- ☐ **Much more** compassionate than the person I want to be

Compared to the <u>fair</u> person I want to be, I am…
- ☐ **Much less** fair than the person I want to be
- ☐ **Moderately less** fair than the person I want to be
- ☐ **Slightly less** fair than the person I want to be
- ☐ **Exactly as** fair as the person I want to be
- ☐ **Slightly more** fair than the person I want to be
- ☐ **Moderately more** fair than the person I want to be

☐ **Much more** fair than the person I want to be

Compared to the <u>generous</u> person I want to be, I am…
☐ **Much less** generous than the person I want to be
☐ **Moderately less** generous than the person I want to be
☐ **Slightly less** generous than the person I want to be
☐ **Exactly as** generous as the person I want to be
☐ **Slightly more** generous than the person I want to be
☐ **Moderately more** generous than the person I want to be
☐ **Much more** generous than the person I want to be

Compared to the <u>moral</u> person I want to be, I am…
☐ **Much less** moral than the person I want to be
☐ **Moderately less** moral than the person I want to be
☐ **Slightly less** moral than the person I want to be
☐ **Exactly as** moral as the person I want to be
☐ **Slightly more** moral than the person I want to be
☐ **Moderately more** moral than the person I want to be
☐ **Much more** moral than the person I want to be

Compared to the <u>ethical</u> person I want to be, I am…
☐ **Much less** ethical than the person I want to be
☐ **Moderately less** ethical than the person I want to be
☐ **Slightly less** ethical than the person I want to be
☐ **Exactly as** ethical as the person I want to be
☐ **Slightly more** ethical than the person I want to be
☐ **Moderately more** ethical than the person I want to be
☐ **Much more** ethical than the person I want to be

Compared to the <u>hard-working</u> person I want to be, I am…

- ☐ **Much less** hard-working than the person I want to be
- ☐ **Moderately less** hard-working than the person I want to be
- ☐ **Slightly less** hard-working than the person I want to be
- ☐ **Exactly as** hard-working as the person I want to be
- ☐ **Slightly more** hard-working than the person I want to be
- ☐ **Moderately more** hard-working than the person I want to be
- ☐ **Much more** hard-working than the person I want to be

Compared to the <u>honest</u> person I want to be, I am…
- ☐ **Much less** honest than the person I want to be
- ☐ **Moderately less** honest than the person I want to be
- ☐ **Slightly less** honest than the person I want to be
- ☐ **Exactly as** honest as the person I want to be
- ☐ **Slightly more** honest than the person I want to be
- ☐ **Moderately more** honest than the person I want to be
- ☐ **Much more** honest than the person I want to be

Compared to the <u>loyal</u> person I want to be, I am…
- ☐ **Much less** loyal than the person I want to be
- ☐ **Moderately less** loyal than the person I want to be
- ☐ **Slightly less** loyal than the person I want to be
- ☐ **Exactly as** loyal as the person I want to be
- ☐ **Slightly more** loyal than the person I want to be
- ☐ **Moderately more** loyal than the person I want to be
- ☐ **Much more** loyal than the person I want to be

Compared to the <u>respectful</u> person I want to be, I am…
- ☐ **Much less** respectful than the person I want to be
- ☐ **Moderately less** respectful than the person I want to be
- ☐ **Slightly less** respectful than the person I want to be

☐ **Exactly as** respectful as the person I want to be
☐ **Slightly more** respectful than the person I want to be
☐ **Moderately more** respectful than the person I want to be
☐ **Much more** respectful than the person I want to be

Compared to the <u>likeable</u> person I want to be, I am…
☐ **Much less** likeable than the person I want to be
☐ **Moderately less** likeable than the person I want to be
☐ **Slightly less** likeable than the person I want to be
☐ **Exactly as** likeable as the person I want to be
☐ **Slightly more** likeable than the person I want to be
☐ **Moderately more** likeable than the person I want to be
☐ **Much more** likeable than the person I want to be

Compared to the <u>warm</u> person I want to be, I am…
☐ **Much less** warm than the person I want to be
☐ **Moderately less** warm than the person I want to be
☐ **Slightly less** warm than the person I want to be
☐ **Exactly as** warm as the person I want to be
☐ **Slightly more** warm than the person I want to be
☐ **Moderately more** warm than the person I want to be
☐ **Much more** warm than the person I want to be

Compared to the <u>friendly</u> person I want to be, I am…
☐ **Much less** friendly than the person I want to be
☐ **Moderately less** friendly than the person I want to be
☐ **Slightly less** friendly than the person I want to be
☐ **Exactly as** friendly as the person I want to be
☐ **Slightly more** friendly than the person I want to be
☐ **Moderately more** friendly than the person I want to be

☐ **Much more** friendly than the person I want to be

Compared to the <u>competent</u> person I want to be, I am…
☐ **Much less** competent than the person I want to be
☐ **Moderately less** competent than the person I want to be
☐ **Slightly less** competent than the person I want to be
☐ **Exactly as** competent as the person I want to be
☐ **Slightly more** competent than the person I want to be
☐ **Moderately more** competent than the person I want to be
☐ **Much more** competent than the person I want to be

Compared to the <u>intelligent</u> person I want to be, I am…
☐ **Much less** intelligent than the person I want to be
☐ **Moderately less** intelligent than the person I want to be
☐ **Slightly less** intelligent than the person I want to be
☐ **Exactly as** intelligent as the person I want to be
☐ **Slightly more** intelligent than the person I want to be
☐ **Moderately more** intelligent than the person I want to be
☐ **Much more** intelligent than the person I want to be

Compared to the <u>skilled</u> person I want to be, I am…
☐ **Much less** skilled than the person I want to be
☐ **Moderately less** skilled than the person I want to be
☐ **Slightly less** skilled than the person I want to be
☐ **Exactly as** skilled as the person I want to be
☐ **Slightly more** skilled than the person I want to be
☐ **Moderately more** skilled than the person I want to be
☐ **Much more** skilled than the person I want to be
Ethical Idealism

You will find a series of general statements listed below. Each represents a commonly held opinion and there are no right or wrong answers. You will probably disagree with some items and agree with others. We are interested in the extent to which you agree or disagree with such matters of opinion. Please each statement carefully. Then indicate the extent to which you agree or disagree.

	Strongly Agree	Moderately Agree	Slightly Agree	Neither Agree nor Disagree	Slightly Disagree	Moderately Disagree	Strongly Disagree
A person should make certain that their political actions associated with their beliefs about multiculturalism never intentionally harm another even to a small degree							
Harming another person with an opposing viewpoint on multiculturalism is wrong, irrespective of how small the harms might be							
Harming others to serve one's beliefs about multiculturalism is always wrong, irrespective of the benefits to be gained							
One should never psychologically or physically harm another person to advance his or her position on multiculturalism							
One should not perform an action which might in any way threaten the dignity and welfare of another individual, even if that individual holds an opposing position on the issue of multiculturalism							
If a political action designed to serve beliefs about multiculturalism could harm an innocent other, then it should not be done							

Deciding whether or not to perform an act by balancing the positive consequences of the act against the negative consequences of the act is immoral, even if the act serves beliefs about multiculturalism							
The dignity and welfare of people should be the most important concern in any society, even for people with opposing positions on the issue of multiculturalism							
It is never necessary to sacrifice the welfare of others, even people with opposing positions on the issue of multiculturalism							

Manipulation Check

In this study you were asked to report the extent to which you would support/oppose enacting various behaviors toward protesters at the upcoming Alt-Right/Antifa rally. What protesters did you evaluate during this task?

☐ Supporters of multiculturalism
☐ Opponents of multiculturalism
☐ Another group of protesters not listed here

Demographics

Please provide us with a bit of information about yourself.

1. Gender:
- Male
- Female

2. Age:

3. Is English your primary language?
- Yes
- No

4. Ethnicity (check all that apply)
☐ African American
☐ Asian or Pacific Islander
☐ Caucasian (White)
☐ Latino
☐ Native American
☐ Other

5. What is the highest level of education you have completed?
- Less than high school
- High school/GED
- Some college
- 2-year college degree
- 4-year college degree
- Master's degree
- Doctoral degree

Attention Check Block

To what extent should we trust your data?
☐ Not at all
☐ Slightly
☐ Moderately
☐ Much
☐ Very much

How distracted were you while completing this study?

☐ Not at all
☐ Slightly
☐ Moderately
☐ Much
☐ Very much

To what extent were you paying attention in this study?
☐ Not at all
☐ Slightly
☐ Moderately
☐ Much
☐ Very much

What is the topic of this study? Sometimes participants do not carefully read the instructions. To correctly answer this question, please select the option "other" and write down the name of your favorite movie.
☐ Sexism
☐ Multiculturalism
☐ I don't remember
☐ Other (please specify:_____)

Do you have any other thoughts about today's study?

Allison B. Mueller
Curriculum Vitae

1007 W. Harrison Street (M/C 285), Chicago, Illinois 60607
Phone: (847) 687-8863 | E-mail: amuelle4@uic.edu

EDUCATION

Ph.D. **University of Illinois at Chicago**, 2019
    Social Psychology
    Minor: Statistics, Methods, and Measurement
M.A. **University of Illinois at Chicago**, 2015
    Social Psychology
B.A. **Northwestern University**, 2009
    Psychology

AWARDS

2019   UIC Cross Program Conference Best Poster Award ($25)
2017-19   UIC Graduate College Student Presenter Award ($300 over three years)
2017-19   UIC Graduate Student Council Travel Award ($825 over three years)
2015   SPSP Graduate Student Travel Award ($500)
2014-15   UIC Chancellor's Graduate Research Fellowship ($8,000 over two summers)
2014-19   UIC Department of Psychology Travel Award ($3,400 over six years)
2014   UIC Department of Psychology Thesis Research Grant ($150)
2008   NU Institute for Policy Research Undergrad Research Fellow Award ($3,150)

PUBLICATIONS

Mueller, A. B., & Skitka, L. J. (in press). Moral courage and moral disengagement: Different sides of the same coin? To be in Newman, L. S. (Ed.), *Why are they doing this to us? The social psychology of genocide and extreme intergroup violence*. Oxford University Press.

Skitka, L. J., Hanson, B. E., Washburn, A. N., & **Mueller, A. B.** (2018). Distinguishing between moral and religious conviction. *PlosONE*.

Washburn, A. N., Hanson, B. E., Motyl, M., Skitka, L. J., Yantis, C., Wong, K. M., Sun, J., Prims, J. P., **Mueller, A. B.**, Melton, Z. J., Carsel, T. S. (2018). Why do some psychology researchers resist using proposed best research practices? A description of perceived barriers. *Advances in Methods and Practices in Psychological Science*, *1*, 166-173.

Mueller, A. B., & Skitka, L. J. (2017). Liars, damned liars, and zealots: The effect of moral conviction on transgressive advocacy acceptance. *Social Psychological and Personality Science*, *9*(6), 711-718.

Motyl, M., Demos, A., Carsel, C., Hanson, B., Melton, Z., **Mueller, A. B.**, Prims, J.P., Washburn, A. N., Wong, K., Yantis, C., & Skitka, L. J. (2017). The state of social and personality science: Rotten to the core, not so bad, getting better, or getting worse? *Journal of Personality and Social Psychology*, *113*(1), 34-58.

Victorson, D., Peterman, A., Bode, R., Buono, S., **Mueller, A.**, Moy, C., & Cella, D. (2015). Development and clinical validation of a new item bank and short form of emotional and behavioral dyscontrol for major neurological disorders: Results from the Neuro-QOL study. *Journal of Neurological Disorders and Stroke*, *3*(2), 1098.

Carlozzi, N., Beaumont, J. L., Victorson, D., Sung, V., Cheng, W., Gorin, B., Sheng Duh, M., Samuelson, D., Tulsky, D., Gutierrez, S., Nowinski, C., **Mueller, A. B.**, Shen, V., & Frank, S. (2014). HD-PRO-TRIAD™ validation: A patient-reported instrument for the symptom triad of Huntington Disease. *Tremor and Other Hyperkinetic Movements*, *4*.

Victorson, D., Carlozzi, N., Frank, S., Beaumont, J. L., Cheng, W., Gorin, B., Sheng Duh, M., Samuelson, D., Tulsky, D., Gutierrez, S., Nowinski, C., **Mueller, A. B.**, Shen, V., & Sung, V. (2014). Identifying motor, emotional-behavioral, and cognitive deficits that comprise the triad of HD symptoms from the patient, caregiver, and provider perspectives. *Tremor and Other Hyperkinetic Movements*, *4*.

Victorson, D., Cavasos, J., Holmes, G., Reder, A., Wojna, V., Nowinski, C., Miller, D., Buono, S., **Mueller, A. B**., Moy, C., Cella, D. (2014). Validity of the Neurology Quality of Life (Neuro-QoL) measurement system in adult epilepsy. *Epilepsy & Behavior*, *31*, 77-84.

TECHNICAL REPORT

Carsel, T. S., Hanson, B. E., & **Mueller, A. B.*** (2017). Perceptions of non-academic internships. Manuscript prepared for the Executive Committee of the Society for Personality and Social Psychology. **[*All authors contributed equally to this paper]**

MANUSCRIPTS IN REVISION

Skitka, L. J., **Mueller, A. B.**, Wei, K. Y., & Melton, Z. (R&R). The gender gap: Who is (and is not) included on graduate level syllabi in social/personality psychology. *Personality and Social Psychology Bulletin.*

Mueller, A. B. (invited resubmission, *PSPR*). A self-regulatory theory of moral licensing.

ORAL PRESENTATIONS

Mueller, A. B., & Demos, A. P. (2019, February). *Revisiting the relative replicability of social/personality and cognitive psychology.* Paper presented at the annual meeting of the Society for Personality and Social Psychology. Portland, OR.

Mueller, A. B., & Demos, A. P. (2018, April). *Revisiting the relative replicability of social/personality and cognitive psychology.* Paper presented at the annual meeting of the Midwestern Psychological Association. Chicago, IL.

Prims. J. P., Demos, A., **Mueller, A. B.** (2018, April). *Institutional prestige: A help or hindrance to large samples?* Paper presented at the annual meeting of the Midwestern Psychological Association. Chicago, IL.

Washburn, A. N., Hanson, B. E., Motyl, M., Skitka, L. J., Yantis, C., Wong, K. M., Sun, J., Prims, J. P., **Mueller, A. B.**, Melton, Z. J., Carsel, T. S. (2018, April). *Why do some psychology researchers resist using proposed best research practices? A description of perceived barriers.* Paper presented at the annual meeting of the Midwestern Psychological Association. Chicago, IL.

Mueller, A. B., & Skitka, L. J. (2018, March). *Liars, damned liars, and zealots: The effect of moral mandates on transgressive advocacy acceptance.* Paper presented at the annual meeting of the Society for Personality and Social Psychology. Atlanta, Georgia.

Mueller, A. B., & Skitka, L. J. (2016, May). *The moralization of personality traits.* Paper presented at the annual meeting of the Midwestern Psychological Association. Chicago, IL.

Mueller, A. B., & Skitka, L. J. (2015, November). *The moralization of personality traits.* Paper presented at the Moral Research Lab Conference on Moral Judgment and Decision Making at the University of Chicago. Chicago, IL.

Mueller, A. B., & Skitka, L. J. (2015, April). *Licensed to lie: Evidence for licensing in persuasive communication.* Paper presented at the annual meeting of the Midwestern Psychological Association. Chicago, IL.

Mueller, A. B., & Skitka, L. J. (2013, March). *Licensed to lie: The moderating role of moral conviction in the link between moral self-licensing and lying.* Paper presented at the 5th meeting of Chicagoland Morality Researchers (C-MORE). Chicago, IL.

POSTER PRESENTATIONS

Carsel, T. S., Hanson, B. E., & **Mueller, A. B.*** (2017, January). *Perceptions of non-academic internships.* Poster presented at the annual meeting of the Society for Personality and Social Psychology. San Antonio, TX. **[*All authors contributed equally to this presentation]**

Mueller, A. B. (2017, January). *A self-regulatory framework of moral self-licensing.* Poster presented at the annual meeting of the Society for Personality and Social Psychology. San Antonio, TX.

Mueller, A. B., & Skitka, L. J. (2016, January). *The moralization of personality traits.* Poster presented at the annual meeting of the Society for Personality and Social Psychology. San Diego, CA.

Mueller, A. B., & Skitka, L. J. (2015, April). *Licensed to lie: Evidence for licensing in persuasive communication contexts.* Poster presented at the first annual UIC Cross-Program Conference. Chicago, IL.

Mueller, A. B., & Skitka, L. J. (2015, April). *Licensed to lie: Evidence for licensing in persuasive communication contexts.* Poster presented at the annual UIC Student Research Forum. Chicago, IL.

Mueller, A. B., & Skitka, L. J. (2015, February). *Licensed to lie: Evidence for licensing in persuasive communication contexts.* Poster presented at the annual meeting of the Society for Personality and Social Psychology. Long Beach, CA.

Mueller, A. B., & Bonam, C. (2014, May). *The effect of respondent race on color-blind ideology endorsement.* Poster presented at the annual meeting of the Midwestern Psychological Association. Chicago, IL.

Mueller, A. B., & Bonam, C. (2014, February). *A post-racial America? The role of perceived racial equality in the link between respondent race and color-blind ideology.* Poster presented at the annual meeting of the Society for Personality and Social Psychology. Austin, TX.

REPRESENTATIVE MEDIA MENTIONS

Daily Mail	*PsychCentral.com*	*Phys.org*
Deseret News	*ScienceDaily.com*	*Brigham Young Radio*

TEACHING EXPERIENCE

Instructor
- Social Psychology (Summer 2017)
- Statistical Methods in Behavioral Sciences (Spring 2017)

Mentor, Illinois Math and Science Academy: Student Inquiry & Research Program
- Mentored an advanced high school student for one academic year on all aspects of the research process (August 2015 – April 2016)

Guest Lecturer
- Psychology of Women and Gender
  - Psychology and Women's Health Issues (October 24, 2017)
- Media and Professional Writing (offered through UIC's English Department)
  - Liars, Damned Liars, and Zealots: Moral Mandates and Journalistic Integrity (October 5, 2017)
- Social Psychology
  - Social Psychological Research Methods (August 31, 2017)
  - Aggression (November 10, 2016)
- Statistical Methods in Behavioral Sciences
  - Dependent Means t-Tests (March 9, 2017)
  - Independent Means t-Tests (November 3, 2016)
- Introduction to Research in Psychology
  - Sampling, Descriptive Statistics, and Data Displays (April 7, 2014)

Teaching Assistant
- All courses related to Social Psychology (2012 – 2019)

Participant in Teaching Training Programs
- UIC Practicum in Teaching (Fall 2016 – Spring 2017)
- SPSP Mentor Lunch: Careers at Liberal Arts Schools (January 2017)

RESEARCH EXPERIENCE

Senior Web Testing & Marketing Analyst (April 2019 – Present)
- Quill.com
  - Manage analytical processes and A/B testing pipeline to improve the profitability and efficiency of eCommerce marketing programs
    - Own test design, test implementation, data analysis, presentation of results, and recommendations

Quantitative Research Intern (May 2018 – August 2018)
- C+R Research
  - Co-designed and implemented surveys; assured data quality; built client reports

Research Consultant (2018 – 2019)
- Society for Personality and Social Psychology (SPSP)
  - Assessed SPSP members' perceptions of diversity, climate, and sexual harassment issues within society conventions and programs

Research Consultant (2016 – 2017)
- Society for Personality and Social Psychology (SPSP)
  - Assessed SPSP members' interest in and support for graduate student participation in non-academic internships
    - Designed and implemented three surveys for SPSP students, faculty, and industry partners, respectively
    - Analyzed the data from these surveys using both quantitative and qualitative methods
    - Co-authored an executive report, a short report, and a PowerPoint presentation to summarize the findings
  - Compiled a database ($N = 60$) of non-academic internships currently available to SPSP graduate students

Research Consultant (2014 – 2016)
- Ministry Leadership Center, funded by the Hilton Foundation
  - Observed and identified themes from five meetings about the formation process for leaders in Catholic institutions, such as Catholic health care systems and Catholic higher education

Research Study Coordinator (2009 – 2012)
- Northwestern University Department of Medical Social Sciences
  - Coordinated several clinical trials aimed at improving quality of life assessment for people with advanced cancer, Huntington's disease, and ALS

AFFILIATIONS AND MEMBERSHIPS

- Association for Psychological Science
- Midwestern Psychological Association
- Society for Personality and Social Psychology
- The Society for the Psychological Study of Social Issues

SERVICE TO THE PROFESSION

- Ad hoc journal reviewer:
    - *Personality and Social Psychology Review* (2018)
    - *Psychological Science* (2017)
    - *Journal of Personality and Social Psychology* (2016)
    - *European Journal of Social Psychology* (2016)
    - *Journal of Psycho-Oncology* (2010)
- Volunteer, annual meeting of the International Society of Political Psychology (2012)

DEPARTMENTAL SERVICE

- Social/Personality Division Teaching Assistant
    - Coordinated Prospective Graduate Student Visiting Day, Brown Bag lunches
- Member, UIC Cross Program Conference Planning Committee (2015 – 2016)
- Member, UIC Psychology Diversity Advancement Committee (2013 – 2019)