# MORAL HUMILITY: AN ANTIDOTE TO THE DARK SIDE OF MORALITY

By

Shree Vallabha

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology — Doctor of Philosophy

2025

**ABSTRACT**

Morality has a dark side. Our moral tendencies breed rigidity, conflict, extremism, hate, intolerance, and violence. In this project, I proposed moral humility as one antidote to these dark features of our morality. Given that moral humility is a largely empirically unexplored phenomenon, across ten studies ($N$ = 10, 978) in this project, I investigated its measurement and structure, its nature and correlates, and its implications in a context characterized by the dark features of our morality.

To these ends, I first developed a moral humility scale using psychometric factor analytic methods (EFA and CFA), which comprised of three factors—moral fallibility, moral openness/learning, and moral superiority. Second, I probed its nomological network, specifically testing its relationship to personality, religiosity, ideology, political extremity, moral grandstanding, and other relevant constructs. Moral humility was associated with other constructs in expected and sensible ways, such as higher openness to experience and modesty, and lower moral grandstanding and political extremity. Third and importantly, I tested its predictive validity in the context of political polarization in the US. Polarization was chosen as the context because research has suggested that it is a moralized context with features associated with the dark aspects of morality. Across correlational studies, I found that moral humility was associated with lower levels of polarization across a range of outcomes. It was associated with lower antipathy and antagonism towards outgroup, lower rigidity in one's own political views, and lower rejection of compromise and contact, amongst other outcomes. These findings provided strong evidence to suggest that moral humility could be ameliorative in contexts that bring out our dark moral nature. These studies also established moral humility's incremental validity over related and important constructs like moral relativism and intellectual humility, indicating that moral humility is likely a distinct construct and has unique explanatory value.

Finally, I designed five moral humility interventions and tested its impact on moral humility and polarization using experimental design. I found that some interventions worked more consistently in increasing moral humility – providing evidence that moral humility can be targeted and increased using interventions. These interventions also lowered polarization, such as decreasing antipathy against political outgroup and increasing willingness toward cross-cutting exposure, thus further supporting moral humility's predictive and causal role in moralized and high-conflict situations. The experimental studies also established moral humility's discriminant validity from close constructs like moral relativism and intellectual humility, reinforcing that moral humility offers a distinct and meaningful mechanism through which moralized conflicts and contexts can be understood and

addressed. This project also highlighted gaps in the current understanding of moral humility and laid the groundwork for future inquiry into the nature and significance, such as its impact on moral behaviors and moral improvement, mechanism underlying change and development of moral humility, and contextual and cultural variations in moral humility.

Taken together, this project is the first comprehensive investigation into moral humility. By developing a validated measure, mapping its conceptual network, and demonstrating its predictive and causal role in moralized context like political polarization, this research establishes moral humility as a theoretically meaningful and practically consequential construct. It provides insight into how moral humility could help counter the dark and ugly aspects of our morality and quell the rigidity, intolerance, and antagonism that often accompany moralized intergroup conflicts. Finally, it opens doors for a deeper understanding of moral humility across outcomes, individuals, context, and cultures.

Dedicated to the realization of
*Saccidānanda*

**ACKNOWLEDGEMENTS**

With this dissertation, a very consequential and meaningful chapter of my life comes to a close. I want to take a moment to acknowledge all those who have profoundly enriched my life and journey so far, and have especially made it possible for me to reach this significant milestone.

Foremost, my advisor, Mark Brandt, who has poured *so much* energy, time, and resources into transforming me into a careful and rigorous scholar, nothing I will say here will be enough to fully express my gratitude to him. For whatever reason, he took a chance on me and my potential – someone from thousands of miles away, from a different culture and educational context – and allowed me the privilege of learning how to do good research from one of the best in the field. For five years, a very substantial chunk of his life and mine – he has generously, earnestly, and to a great extent selflessly, invested in my development as a social scientist, a psychologist, a researcher, a theorist, a methodologist, an academic, a teacher, a mentor, a public speaker, a writer, a professional, and a person with a meaningful and rewarding life. Very few people have an inordinate impact on your life—Mark is one of those rare people whose impact on my life is and will always be immeasurable and life-changing. Everything that I have accomplished so far academically and will accomplish academically in the future, I owe and dedicate to him, and even that feels insignificant compared all that he has done for me. I remember him telling me during my first semester as his mentee that he wanted to mentor me in a way that would prepare me to stand confidently as an independent scholar by the end of my PhD. I believe that not only has he nurtured me into becoming that but also very importantly exemplified for me the value of being an ethical, kind, and responsible scholar. I will always be grateful for the privilege of having been his student. I hope to take all that I have learnt from him into the next phase of my life, and carry forward at least some of his extraordinary creativity, rigor, prolificacy, integrity, generosity, and goodness through my own future pursuits. I look forward to many more years of collaboration and I hope I will make him proud.

I am also immensely thankful to other outstanding mentors and academics whom I have known closely and who have enriched my intellectual and academic development. This includes Rich Lucas, Jennifer Wolak, and Bill Chopik — who I am fortunate to have as my dissertation committee. Taking Rich's class in grad school and engaging with his intellectual process through brownbags and meetings fundamentally changed my outlook to research and teaching. It has moved me to be attracted to the big picture in science and to be ambitious in pushing the boundaries of your field. Jenny's class was the first one that I took in grad school that spoke to my own

substantive interests. It was in this class, that through her very invested and systematic mentorship, I “birthed” the idea and research on moral humility, and simultaneously learnt how to go about methodically formulating research questions in psychology. I am deeply grateful for that. I never formally had a secondary advisor during graduate school, but I felt that Bill always had my back and that someone apart from my advisor was always looking out for my best interests. I am always in awe of his creative thinking as a scientist, and proactive and supportive spirit as a mentor—qualities that I hope to emulate in my own academic career. I am grateful to late Debby Kashy for making me love statistics again. I am also very grateful to my mentors from my previous academic institution, CBCS, especially Prof Narayanan Srinivasan who embodies extraordinary scientific excellence and cultivated in me deep intellectual curiosity and critical thinking — which forms the bedrock of academic identity. Without his and Prof Bhoomika Kar’s support and encouragement, I could not have gotten here.

My family is the core of my meaning, existence, and bliss. They have always been a reminder to me of the beauty and privilege of life, which sometimes got obscured by the stress and isolation of graduate school. All my pursuits, intellectual and otherwise, spring from the strength and meaning I derive from the immensity of their love, support, and wisdom. I am eternally thankful to them for everything and could never get to where I am today without them. My special thanks a few family members. My parents who have been unimaginably and unconditionally supportive of me through my whole life, and especially when I chose a field that wasn’t very respected in their social milieu, and when I further wanted to go to a faraway land to pursue it. Honestly, thanking them using mere words is the hardest and feels the most inadequate. I will spend my whole life thanking them through my actions and devotion. My little brother, Mukunda, who has been my best friend and always challenged my worldviews, routines, and habits — pushing me to continuously grow and never become stagnant or dogmatic. Amruta, Omkar, and the kids, for being a home away from home; I found so much warmth in your abode. And finally, the love of my life, my husband Pranav, who first inspired me to listen to my intellectual instincts over ten years ago, who through his extraordinary philosophical aptitude helped me develop clearer conceptual thinking about my research during graduate school, who has elevated my life spiritually and aesthetically, who fills my life with so much hope, joy, love, courage, and adventure that I feel like a hero(ine) in a great story, and with whom I see a future so thrilling that I am eager to leap forward into the next chapter of life.

My immense gratitude to all my friends, colleagues, and lab members who have so beautifully filled the interstices of these last five years that I shudder to imagine a life without them; the picture would have been indeed

very dreary if they were not in there. I especially appreciate all the wonderful, smart, and kind graduate students I had the privilege of interacting and/or working with. Special thanks to my lab sister Abby for being such a big part of my time here in USA. I have been so fortunate to have met her, worked with her, laughed with her, travelled with her, been vulnerable with her, discussed life, research, and politics with her—it has truly been wholesome. I have learnt so much from her and deeply admire her for her warmth, resilience, and intellect. I look forward to our life-long friendship and collaborations. Shoutout to Kenya and Hyewon who have been very sharp, interesting, and amazing friends and colleagues. I have had such a heartwarming and intellectually enriching time with them, I could not have asked for better companions on this shared journey. Thanks to Jeewon for being the first colleague to make Lansing feel warm and welcoming, and for being absolutely inspirational. Thanks to Brian and Lindsay for often being a social glue in our program and hosting wonderful get togethers (and allowing me the joy of being around their beautiful kids). Thanks to Rebekka and Mariah for their help and guidance during the job market. Thanks also to all the undergraduate research assistants and students I worked with during my time in graduate school—for improving me and my research.

Finally, thanks to the Psychology Department at Michigan State University for supporting me and providing me with multiple opportunities to learn and grow academically. Thanks also to Michigan State University and USA for a life-changing and memorable experience. I will definitely look back at these five years as one of the best times of my life.

**TABLE OF CONTENTS**

## Chapter I: Introduction

Morality plays a unique, central, and multifaceted role in human psychological experience. Our moral tendencies motivate noble and inspiring acts of heroism, altruism, kindness, sacrifice, cooperation, and collective action (Curry et al., 2019; Haidt, 2012; Tomasello & Vaish, 2013; Van Zomeren et al., 2012). For these reasons, people have always considered our moral sense as one of our chief human sensibilities. This sentiment is echoed in Charles Darwin's writing on human morality (1871), "I fully subscribe to the judgment of those writers who maintain that of all the differences between man and the lower animals, the moral sense or conscience is by far the most important …it is summed up in that short but imperious word ought, so full of high significance. It is the most noble of all the attributes of man."

However, morality is a double-edged sword. Apart from its noble side, our capacity for morality also has a dark side. Decades of psychological work has uncovered how our moral tendencies breed rigidity, conflict, hate, intolerance, and violence (Baumeister, 1999; Fiske & Rai, 2014; Kovacheff et al., 2018; Skitka et al., 2021; Pretus et al., 2023). This is because when people process something as a moral concern, i.e., as a matter of right and wrong, or good and bad, people consider their moral stance on the matter as objective and universal. They further imbue their moral stance with intense emotions and prescriptive motivations. Consequently, they become rigid, extreme, intolerant of disagreement, and reject compromise (Garrett, 2019; Ryan, 2014; Yoder & Decety, 2022; Clifford, 2019; Delton et al., 2020; Goodwin & Darley, 2008; Jung & Clifford, 2024; Kodapanakkal et al., 2022; Pretus et al., 2023; Ryan, 2017; Skitka et al., 2021; Van Bavel et al., 2012). In other words, their thinking and approach becomes black-and-white. People consider themselves to be on the "good" side and those who think or act differently are perceived as bad or evil. People become willing to adopt violent solutions to conflicts and use morality to legitimize such actions and tendencies (Baumeister, 1999; Fiske & Rai, 2014; Haidt, 2012; Pinker, 2008; Skitka et al., 2021). Indeed, many perpetrators of violence and atrocities are found to motivated by morality, i.e., doing what is morally good (Fiske & Rai, 2014). They believe their violent and oppressive actions are justifiable and necessary means to a noble, moral end (Baumeister, 1999; Fiske & Rai, 2014; Skitka & Mullen, 2002; Effron & Miller, 2012; Haidt, 2012; Reicher et al., 2008; Sedikides et al., 2014; Kovacheff et al., 2018). Our moral sense therefore "has the nasty habit of always putting the self on the side of the angels." (Pinker, 2008)

Considering how our morality traps us into a rigid mindset, making us quick to demonize others, and paradoxically, blinds us to our own potentially harmful tendencies, I suggest that it might be fruitful to find

countervailing forces that can counter the dark side of our morality such as moral disdain, moral righteousness, moral supremacy, and moral rigidity. In the current project, building on the psychological literature on the virtue of humility (McElroy et al., 2019; Wright et al., 2017; Davis at el., 2016; Van Tongeren et al., 2019), I advance a form of humility, moral humility, as one such antidote to our morally dark tendencies. Thus, in this project, I first explore the idea of moral humility and why it's a suitable antidote to the dark aspects of our morality. Second, I develop a psychometrically valid measure for it. Third, I test moral humility's association with other constructs. Fourth, I test if moral humility has positive implications in the context of a present-day moral conflict i.e., political polarization in the USA. Fifth, I develop and validate interventions of moral humility, test if it can be changed, and examine if it has downstream outcomes on conflicts like political polarization.

**Humility**

Humility is understood as an attribute or quality that lowers self-focus and increases other-focus (Wright et al., 2017). That is, it "involves a shift from the narrow preoccupation with self…into the broader consideration of self *and* other" (Wright et al., 2017, p. 4). Specifically, having humility has been associated with having an accurate view of the self, which involves attributes like acknowledging one's limitations and weaknesses. It is also associated with having open-mindedness and flexibility which involves attributes like a general desire to learn and correct mistakes. And with having an other-oriented interpersonal stance which involves attributes like restraint of egotism, being modest in self-presentation, and a respectful attitude towards others' ideas, skills, and abilities (McElroy et al., 2019; Wright et al., 2017; Davis at el., 2016; Van Tongeren et al., 2019). A humble person would be someone who understands that they are not perfect, that they have scope to learn and improve, and holds an attitude of respect and openness, and not superiority and disdain towards others.

The work on virtues has highlighted that humility occurs in different types of domains or contexts and can accordingly vary across these. For example, a politician might be high in humility about their athletic abilities, but low in humility about their moral traits, or an engineer might be high in humility about their social skills, but low in humility about their technical abilities. Thus, the construct of humility may be too general. For these reasons, identifying important domains which can make the idea of humility more specific and increase its predictive efficacy has been considered fruitful (Davis at el., 2016; McElroy et al., 2014; Hoyle et al., 2016; Leary et al., 2017; Ballantyne, 2023; Van Tongeren et al., 2019).

## Moral Humility

In line with this, I propose that morality may be an important domain for enriching our understanding of humility. Thus, humility in the domain of morality, or what I call moral humility, forms the center of the investigation in this project. The reason why I consider morality an important domain for understanding humility is multifaceted. Before going into these reasons, I first conceptualize moral humility.

### *Definition*

Building on the idea of humility and initial work on moral humility, I conceptualize moral humility as the awareness of one's moral limitations or fallibility, having moral openness or moral learning orientation, and an other-focused moral orientation (Owens et al., 2019, Smith & Kouchaki, 2018; Vallabha et al., 2024). Like humility, it contains both self-oriented and other-oriented dimensions of fallibility, learning, openness, and respect, but specifically about our own and others' *morality.* Thus, a morally humble person acknowledges their moral imperfections, strives towards moral improvement and learning, acknowledges others' moral strengths, and expresses understanding and openness towards moral differences.

### *Importance of the moral domain*

Why should we study humility in the moral domain? Like previously mentioned, morality plays a unique and central role in human psychological experience. For instance, people experience morality as a basic psychological need (Prentice et al., 2019) — they have a fundamental need to feel moral. They consider moral aspects of the self as most central to personal identity (Strohminger & Nichols 2014; Stanley et al., 2020), and moral self-identity predicts well-being and meaning in life (Goering et al., 2024). Other people and groups are judged primarily on their morality, which is associated with how people interact and affiliate with them (Brambilla et al., 2021; Goodwin et al., 2014; Nicolas et al., 2022; Leach et al., 2015; Halevy et al., 2015). Morality is associated with intense and specific emotions (Skitka et al., 2005; Schein & Gray, 2018; Brady et al., 2017; Garrett, 2019; Ryan, 2014; Yoder & Decety, 2022), and moralized attitudes and judgements are more rigid and extreme than non-moral ones (Luttrell, et al., 2016; Van Bavel et al., 2012). Morality binds (us in groups, like religious, political) and blinds us (to the humanity of moral outgroups; Haidt & Kesebir, 2010; Haidt, 2012). Thus, morality is central to our ego and self-concept, making our emotions run hot, our minds closed, and serves as the basis of social affiliation, cooperation, exclusion, and dehumanization. Together this suggests that the moral domain is at the front and center

of our social psychological experience, influencing our cognitions, emotions, and behaviors. These self and other concerning aspects together make morality a relevant and important domain for the study of a virtue like humility.

The most important reason for studying moral humility is that if we look at the dark aspects of our moral nature, like the ones highlighted before, we will notice that moral humility is a virtue that is very well positioned to counter the negative aspects of morality. If morality traps us into rigid and closed mindsets of moral correctness, moral humility opens us to moral growth and learning; if morality traps us into thinking that we are always on the side of angels and makes a villain of those we disapprove, then moral humility enables us to see the other side with respect, as morally worthy and valuable; if morality makes us susceptible to moral superiority, righteousness, and sanctimony, then moral humility helps us become aware of our own moral flaws and weaknesses.

From a measurement perspective, domain-specific psychological traits are valuable as they are narrower in scope and help enhance fidelity and criterion-related validity. However, too much narrowness can come at the cost of predictive bandwidth and tautological tests with criteria (Salgado, 2017). In my view, morality is a domain that is placed well between too much specificity or generality, making the study of moral humility as a domain specific form of humility reasonable. It is suited to narrowly predict criteria that pertains to morality, but at the same time, given the relatively wide range of things the moral domain touches (e.g., politics, religion, cultural norms, leadership, relationships, conflicts), it is well positioned to predict a broader range of outcomes.

***Moral humility in religion, culture, and literature***

Does the idea of moral humility have any grounding in human experience and common understanding? We can indeed find the idea of moral humility contained in various religions and cultures, scholarly and literary works, and public discourse, highlighting its practical importance to humans. For instance, in Christianity, one is advised to reflect and work on one's own moral flaws before being quick in moral condemnation of others.

*"Why do you see the speck that is in your brother's eye, but do not notice the log that is in your own eye? Or how can you say to your brother, 'Let me take the speck out of your eye,' when there is the log in your own eye? You hypocrite, first take the log out of your own eye, and then you will see clearly to take the speck out of your brother's eye."*

*- Matthew 7:3-5, The New Testament*

*"Let he who is without sin cast the first stone."*
*- John 8:7, The New Testament*

*"Judge not, and ye shall not be judged. Condemn not, and ye shall not be condemned. Forgive, and ye shall be forgiven."*
*-Luke 6:37, The New Testament*

A similar advice about searching for darkness within us before finding it in others is found in Indian teachings.

*"Bura Jo Dekhan Main Chala, Bura Na Milya Koye*
*Jo Munn Khoja Apna, To Mujhse Bura Na Koye"*
*Translation: "I searched for evil, but didn't find evil anywhere*
*When I searched myself, I found the biggest evil within."*
*– Kabir's Dohe*

Carl Jung in his works grappled with the dark side within people (also called "shadow" in his work) and wrote about the need for humans to realize their dark aspects to achieve self-actualization. He called it a moral problem requiring considerable moral effort.

*"Unfortunately, there can be no doubt that man is, on the whole, less good than he imagines himself or wants to be. Everyone carries a shadow, and the less it is embodied in the individual's conscious life, the blacker and denser it is. At all counts, it forms an unconscious snag, thwarting our most well-meant intentions."*
*– Carl Jung, Psychology and Religion: West and East*

*"The shadow is a moral problem that challenges the whole ego-personality, for no one can become conscious of the shadow without considerable moral effort. To become conscious of it involves recognizing the dark aspects of the personality as present and real. This act is the essential condition for any kind of self-knowledge.*
*– Carl Jung, Aion*

The need for acknowledging the dark side within us and being alert to our moral weaknesses and moral limitations is also an important theme in literature, along with the theme of recognizing the moral complexity of others.

*"The world isn't split into good people and Death Eaters. We've all got both light and dark inside us."*
*- J.K. Rowling, Harry Potter and The Order of the Phoenix*

*"Many that live deserve death. And some that die deserve life. Can you give it to them? Then do not be too eager to deal out death in judgment. For even the very wise cannot see all ends."*
*-J.R.R. Tolkien, The Lord of the Rings: The Fellowship of the Ring*

A similar theme of moral complexity was invoked in the social messaging of civil rights movement leader, Martin Luther King Jr.

*"We must recognize that the evil deed of the enemy-neighbor, the thing that hurts, never quite expresses all that he is. An element of goodness may be found even in our worst enemy …When we look beneath the surface, beneath the impulsive deed, we see within our enemy-neighbor a measure of goodness and know that the viciousness and evilness of his acts are not quite representative of all that he is… there is some good in the worst of us and some evil in the best of us. When we discover this, we are less prone to hate our enemies"*
*-Martin Luther King Jr., Strength to Love*

Finally, recently, Barack Obama, former US president, spoke against our tendency for moral righteousness and condemnation, and encouraged seeing moral strengths in others.

*"This idea of purity and you're never compromised... you should get over that quickly. The world is messy, there are ambiguities. People who do really good stuff, have flaws. People who you are fighting may love their kids and share certain things with you. There is the view that the way for me to make change is to be as judgmental as possible about other people, and that's enough. That's not bringing about change. If all you are doing is casting stones, you are probably not going to get that far".*

*-Barack Obama*

Together, these examples convey how our stories, teachings, and writings urge us towards shining a light on our moral weaknesses and flaws, so that we can grow morally, treat others with grace, and find goodness in them. Thus, moral humility as conceptualized here, is consistent with ideas and teachings about morality embedded in our culture.

**What Moral Humility Is Not?**

***Moral Humility and Moral Relativism***

One construct moral humility might be confused with is moral relativism. Moral relativism is a metaethical viewpoint that entails believing that there are no universal or objective moral truths and what is morally appropriate is dependent on the standards of a culture or group or person (Gowans, 2021). Does being morally humble mean being a moral relativist? I will test this empirically. However, there is a least a conceptual distinction to be made between these ideas. Moral relativism is a stance towards the objective nature of morality in general (do objective moral truths exist), moral humility involves a stance towards the fallibility of one's *own* morality (do *I have* moral limitations).

Thus, a morally humble person can be so without being a moral relativist. Such a person might think, for instance, that objectively true moral positions do exist, but they themself might not have accurate insight into it, or that they can learn from others moral viewpoints to get at the objective truth. Or they might believe that they do have knowledge of the correct moral perspective but fail to translate into appropriate moral behaviors. These possibilities show that being morally humble does not necessitate believing there are no objectively correct moral answers or ways. Thus, I conceptualize moral humility as distinct from moral relativism and adopt this distinction in this project henceforth. *However,* I do concede that moral humility can and perhaps does *psychologically* co-occur with moral relativism in people. For instance, Smith and Kouchaki (2018) suggest that too much moral humility might manifest as moral relativism. Thus, these constructs are not psychologically incompatible, and hence I do *not* argue that they are *entirely* distinct.

***Moral Humility and Intellectual Humility***

Intellectual humility is a type of humility that has recently received a lot of scholarly attention. It is humility in the domain of knowledge or ideas (Krumrei-Mancuso & Rouse, 2016; Leary et al., 2017). That is, it is the recognition of one's *intellectual* or epistemic limitations (Porter et al., 2022). One might wonder, do we need the

moral humility construct to be a separate psychological construct, to be a different type of humility? Or is moral humility just intellectual humility in the moral domain—i.e., humility about moral ideas and knowledge?

There are good reasons to think that there are important conceptual and empirical differences between these two constructs. First, moral humility can go beyond just cognitive elements. For instance, when we make judgements about our own moral imperfections, or when we deem others as morally better than us in some regard, we are sometimes judging actions, emotions, or characters, as opposed to just knowledge and beliefs. Indeed, virtues have been conceptualized to have multiple psychological components such as knowledge, behavior, motivation, and disposition (Fowers et al., 2021).

Second, even when it comes to just knowledge and beliefs, research suggests that moral beliefs have unique psychological signatures compared to non-moral beliefs — they are more laden with intense emotions, are more rigid and extreme, and motivate more intolerance, inflexibility, and desire for punishment (Clifford, 2019; Delton et al., 2020; Goodwin & Darley, 2008; Jung & Clifford, 2024; Kovacheff et al., 2018; Luttrell, et al., 2016; Pretus et al., 2023; Ryan, 2017; Skitka et al., 2021; Skitka & Mullen, 2002; Van Bavel et al., 2012). In a similar vein, morality is very central to people's self-conception and how they interact and organize their social world, making it an important and distinct aspect of psychological experience (Haidt & Kesebir, 2010; Haidt, 2012; Strohminger & Nichols 2014; Stanley et al., 2020; Brambilla et al., 2021; Goodwin et al., 2014; Nicolas et al., 2022; Leach et al., 2015; Halevy et al., 2015). Together, these reasons suggest that moral humility is at least somewhat conceptually and empirically distinct from intellectual humility, and that these differences should drive us to study it as separate from intellectual humility. This is the path taken in this project. However, I do concede at the outset that moral humility has components that can be considered as aspects of intellectual humility. For example, humility about one's moral knowledge and beliefs can be considered intellectual humility about one's moral knowledge. Thus, I do *not* argue that they are entirely distinct.

**What do we know or don't know about moral humility so far?**

Moral humility is thus far a largely empirically unexplored construct. One study investigated the effect of a leader's *perceived* moral humility on followers in an organizational context (Owens et al., 2019). Within two different samples (in China and US), team members or employees reported their leader's moral humility i.e. the extent to which the team leader acknowledged their own moral limitations, were open to moral learning, and recognized moral strengths in other team members. The team members' ethical and prosocial behavior was also

measured (other-report or self-report). The study found that when team leaders were rated higher in moral humility by the followers, the followers were reported to engage in more ethical and prosocial behavior. The authors suggested that this effect was observed because of enhanced moral self-efficacy on the part of team members because of team leader's humility. This study suggested that moral humility can have positive interpersonal outcomes on ethical and prosocial behavior. Further, the study found these effects over and above the effects of general humility of leaders, suggesting that domain-specific forms of humility, like moral humility, can enhance the predictive efficacy of the humility.

However, this study did not test people's own moral humility and what implications this has on different types of outcomes. It also used measures of moral humility that were not psychometrically validated. Vallabha et al., (2024) conducted the first test that we know of where people's own moral humility was measured and used to predict people's level of affective polarization, support for political compromise, and other attitudes towards political outgroups. Across three samples (national and student samples) in the US, participants reported their own levels of moral humility on face-valid measures, as well as a host of political attitudes capturing political polarization. People higher in moral humility were less inclined to be polarized i.e., they expressed lower negative attitudes and behavioral intentions towards political outgroups. Because the current political relations between the partisan groups in the United States (Republican and Democrats) have been characterized by moral disdain, extremity, and aversion (Finkel et al., 2020, 2024; Garrett & Bankert, 2020; Enders & Lupton, 2021; McGarry et al., 2023; Puryear et al., 2023), the authors reasoned that moral humility might help attenuate such tendencies. This provided the first evidence suggesting that people's own moral humility can have positive outcomes in conflictual contexts rooted in moral divisions. However, like Owens et al. (2019), Vallabha et al. (2024), also used measures of moral humility that were not psychometrically validated. Further, this work like others didn't explore moral humility's nomological network, restricting insight into how it is related to other psychological constructs. Finally, previous work also didn't examine if and how moral humility could be changed, and whether it plays a causal role in outcomes of interest. Thus, we have limited information on the nature of moral humility as a psychological construct and its implications.

## The Current Project

In the present project, I address the limitations of previous empirical work on moral humility and extend it further. Specifically, I (i) develop a measure of moral humility using psychometric factor analytic methods (EFA

and CFA), (ii) probe the nomological network of moral humility, such as how it is related to personality, religiosity, ideology, political extremity, and other relevant constructs, (iii) test it's predictive and criteria validity in reducing divisions in a morally relevant context, (iv) test its incremental validity in reducing divisions over other related constructs like moral relativism and intellectual humility, and (v) develop interventions that change moral humility and test its impact on political divisions. Studies addressing points i through iv were conducted before the dissertation committee provided feedback, thus they are reported together below first in the "Chapter II". After the committee provided feedback, I extended this previous work in a causal direction and developed and tested moral humility interventions. This is described after the previously conducted studies are reported and summarized; they are in the "Chapter III".

**Context of the Study: Polarization in USA**

The predictive/criteria validity of moral humility is tested in the context of present-day political polarization in the USA. Although polarization has many meanings, in this project it is used to describe and measure the phenomena of animus or antipathy towards political outgroups, which has significantly increased in the US over the last two decades (Iyengar et al., 2019; Finkel et al., 2020; Mason, 2015; Druckman & Levy, 2022). The rise in partisan antipathy has raised alarm and spurred investigations on understanding its nature and ways to reduce it. The primary reason for choosing this as the context in which I study moral humility is because this type of polarization provides us with a morally relevant context that has the features associated with the dark aspects of our morality described above.

Specifically, this type of polarization has been described as a "quasi-religious phenomenon" (Finkel et al., 2024), where people have affective attachments akin to religious attachments to their political ingroup, believe in the moral righteousness and supremacy of their political side, judge the other side as immoral, and dehumanize political outgroups (Cassese, 2021; Finkel et al., 2020 Finkel et al., 2024; Lees & Cikara, 2020; Martheus et al., 2021; McGarry et al., 2023; Puryear et al., 2023; Tappin & McKay, 2019). Consistent with this, research shows that the more people engaged with politics in moral terms, the greater aversion they expressed toward the political outgroup, rejected political compromise, and punished politicians who engaged in compromise (Garrett & Bankert, 2020; Grubbs et al. 2020; Ryan, 2017). Political animus has also been associated with political extremity (Brandt & Vallabha, 2024), selective and partisan engagement with information (Hobolt et al., 2023; Levy, 2021), support for

political violence (Kalmoe & Mason, 2022) and anti-democratic attitudes and behaviors (Graham & Svolik, 2020; Kingzette et al., 2021; McCoy et al., 2018; Simonovits et al., 2022; McCoy & Somer, 2019).

Thus, political polarization in the US represents a morally fraught conflict that has been associated with threats to social cohesion and democratic commitments and procedures. Accordingly, it thus provides an apt context for the study moral humility. Theoretically, the expectation is that moral humility will show positive outcomes in any moral situation that brings out of the dark aspects of our moral nature. However, here I only test moral humility's role in a polarized and moralized political context as a case study of moral humility's effects.

## Chapter II: Scale Development, Nomological Associations, Predictive Validity

Seven studies were conducted wherein the main aims were to (i) create a moral humility scale, (ii) probe its nomological network, and (iii) test the scale's predictive and criteria, and incremental validity. The studies and results are described in the order of these three broader questions that I aimed to address. First, all the samples that went into addressing these three broad aims are described together. Then factor analysis, nomological associations, and predictive/criteria validity tests are described in three separate sections, each section accompanied by the description of research design, methodology, measures, and results associated with that section. Overall, the aim of the studies was to understand the nature of moral humility, its nomological network, and test its mitigatory role in political polarization.

### Dataset and Participants

The studies were conducted on seven adult American samples, including samples from Prolific, Michigan State University psychology student pool, and YouGov national samples. The details of the different samples are in Table 1. Prolific is an online service that facilitates the crowdsourcing of research participants (Douglas et al., 2023; Peer et al., 2022). YouGov samples were collected as part of national surveys conducted by the Polarization Research Lab (Sample 4) and CCES (Cooperative Election Study, Sample 6).

Participants were paid $0.85 for doing the study (~5 minutes) in Sample 1, $1.68 in Sample 2 (~10 minutes), $1.68 in Sample 3 (~10 minutes), 0.50 research credits in Sample 5 (~20 minutes), and $2.20 in Sample 7 (~15 minutes). In the Prolific samples (Samples 1, 2, 3, 7), participants were recruited evenly from those who self-identified as Democrats and Republicans in Prolific' s prescreening to circumvent the problem of samples collected online being liberally biased.

Sample size justifications for Sample 3-7 are in "Predicting Polarization Outcomes" section later as the sample size for those samples were determined based on estimates of relationship with political outcomes. Sample 1 was collected primarily for EFA and Sample 2 for CFA and exploring nomological associations. In general, for EFA and CFA, larger samples are better, and all the samples are > 500. A sample size of 800 in Sample 1 gave between a very good to excellent sample size for an EFA (Comrey & Lee, 1992). 800 also gives about a 17:1 sample to item ratio which is above the 15:1 recommended as best practices (Pett et al., 2003). Along similar lines, a sample size of 500 for CFA in Sample 2 also gave a very good sample size for factor analysis (Comrey & Lee, 1992) as well as an 80% power to detect small correlations of $r \sim .1$, useful for probing nomological associations.

**Table 1**

*Sample characteristics*

| Sample Number | Sample Platform | Scale Used | *N* | $M_{age}$ | $SD_{age}$ | % Men | % Women | % White | % Black | % Other Ethnic |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Prolific | Full | 823 | 24.43 | 13.97 | 46.9 | 46.29 | 73.39 | 9.6 | 17.01 |
| 2 | Prolific | Full | 518 | 26.84 | 14.5 | 44.5 | 48.26 | 76.83 | 6.17 | 17 |
| 3 | Prolific | Full | 1499 | 44.94 | 14.2 | 49.03 | 48.76 | 72.18 | 10.14 | 17.68 |
| 4 | YouGov | Short | 1000 | 49.67 | 17.33 | 46.4 | 53.6 | 64.9 | 10.8 | 24.3 |
| 5 | Student | Full | 930 | 19.52 | 1.5 | 24.01 | 72.95 | 65.99 | 7.06 | 26.95 |
| 6 | YouGov | Short | 1000 | 50.97 | 17.56 | 44 | 55.2 | 66 | 14 | 20 |
| 7 | Prolific | Full | 805 | 23.13 | 13.07 | 50 | 47.64 | 74.32 | 7.94 | 17.74 |

**Transparency and Openness**

Sample 4 and 6's study design, hypotheses, and planned analyses were preregistered and can be found at OSF. The data, code, materials, and SOM for all studies is also at OSF. All studies were approved by the Michigan State University Institutional Review Board.

## Scale Development: Exploratory Factor Analysis

**Measures**

***Moral Humility***

Exploratory factor analysis was conducted in Sample 1. 45 items capturing various aspects of moral humility such as moral fallibility, moral limitations, moral openness, moral learning, moral disdain, and moral righteousness were constructed. To this end, items from other humility scales (e.g., McLaughlin et al., 2023) were adapted, items from previous moral humility work (Vallabha et al., 2024) borrowed, and new items generated based on previous theoretical work (e.g., Smith & Kouchaki, 2018). The full set of items are included in Table 2.

Participants were instructed: "Take a moment and think about your strongly held moral beliefs and values that help you navigate moral issues and concerns in everyday life and in society. When you think about your moral views, ideas, and values, please answer the following items about how you see yourself and other people."[1] They

[1] About half the participants received some additional instructions telling them about various domains and aspects of morality people care about (see SOM) before they were saw these instructions. This was done as we were testing if the two sets of instructions would yield different factor solutions, but after observing that they yielded similar factor solutions, we proceeded to treat and analyze them as one measure.

then indicated their response on the 45 items on a 7-point scale (*1 = Strongly Disagree, 2= Moderately Disagree, 3= Slightly Disagree, 4= Neutral, 5= Slightly Agree, 6= Moderately Agree, 7 = Strongly Agree*).

**Analysis and Results**

In the first step of exploratory factor analysis, the number of factors to extract was determined using a scree plot (Figure 1), parallel analysis, and a priori theoretical considerations of moral humility content. Parallel analysis suggested six factors, but a scree plot showed that there were three or four most distinct factors or components, with other factors or components being less distinct. Additionally, there weren't a priori reasons to expect many (>4) distinct factors. Thus, based on these theoretical and empirical considerations, three and four factors were decided upon to be extracted for further consideration.[2] Three factors accounted for 48% of the total variance, while the fourth factor added only 4% additional variance.

In the second step, three and four factors were extracted using minimum residual factor analysis and rotated with oblimin rotation, thus allowing factors to be correlated. In the third step, the number of items were reduced by identifying items that clearly and strongly loaded onto a single factor. For that purpose, factor loadings less than 0.3 (indicated by blank spaces in Table 2) were removed. Then, how the items performed in both the three and four factor solutions was compared and finally only those items that consistently loaded on the same factor and did not load on multiple factors across the solutions were picked.

This resulted in three coherent factors that were named as Moral Fallibility, Moral Learning/Openness, and Moral Superiority. The fourth factor wasn't coherent as it consisted of mostly smaller loadings (<0.3) or cross-loadings. The items chosen for the three factors using the aforementioned criteria are highlighted in bold in Table 2. The final chosen factors and the respective items are also separately shown in Table 3.

The 30-items that were retained for the Moral Humility scale had an overall reliability of 0.95 and excellent subscale reliabilities (Moral Learning/Openness $\alpha = 0.95$, Moral Fallibility $\alpha = 0.94$, Moral Superiority $\alpha = 0.91$).

[2] I did investigate the results of five and six factor solutions. No coherent themes seemed to emerge from additional factors beyond four. Additionally, the fifth and sixth factors accounted for only 4% and 3% variance, which provided further rationale to proceed with three and four factor solutions for further analysis.

**Figure 1**

*Scree plot of eigen values (Sample 1)*

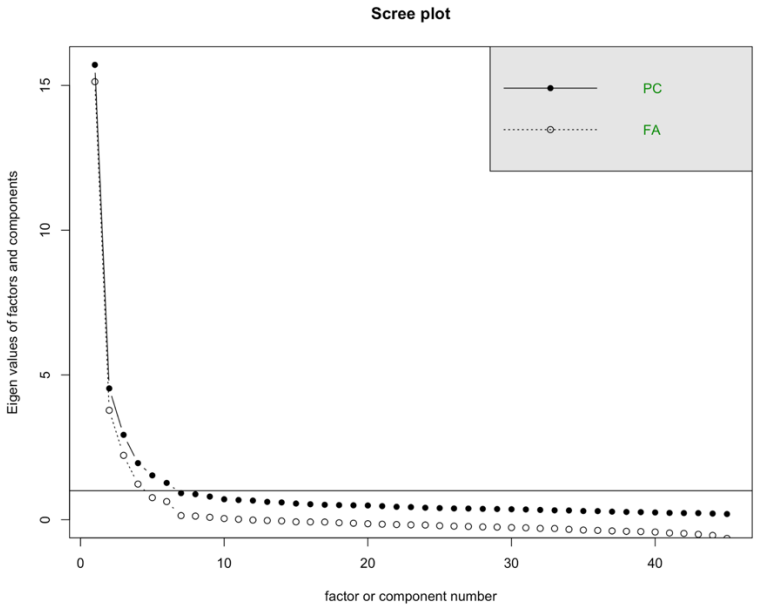



**Table 2**

*Standardized Loadings with Minimum Residual Factor Analysis with Oblimin rotation (Sample 1)*

| | | 3 factors | | | 4 factors | | | |
|---|---|---|---|---|---|---|---|---|
| | | **F1** | **F2** | **F3** | **F1** | **F2** | **F3** | **F4** |
| 1 | I would never change what I believe about moral topics. | | | 0.342 | 0.450 | | | |
| 2 | I doubt I would change my mind on moral issues. | | | 0.379 | 0.535 | | | |
| 3 | I don't think there is anything that could happen that would shift how I think about moral issues. | | | 0.392 | 0.502 | | | |
| 4 | My mind is settled about moral issues. | 0.305 | | 0.412 | 0.506 | | 0.324 | |
| **5** | **My beliefs about moral issues may be incorrect.** ϕ | **0.732** | | | **0.756** | | | |
| 6 | I would be willing to revise what I believe about moral topics. | 0.428 | 0.470 | | 0.598 | 0.346 | | |
| **7** | **I realize that my perspective about moral issues may be wrong.** ϕ | **0.692** | | | **0.760** | | | |
| **8** | **I understand that my moral beliefs about these statements may be limited.** ϕ | **0.578** | | | **0.464** | | | |
| **9** | **I'd be excited to learn from others about moral issues.*** | | **0.703** | | | **0.761** | | |

Table 2 (cont'd)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **10** | **I enjoy hearing diverse perspectives about moral issues. *** | | **0.719** | | | **0.740** | |
| **11** | **I am interested to learn how other people think about moral issues. *** | | **0.663** | | | **0.742** | |
| **12** | **I am open to exploring moral topics more in the future. *** | | **0.662** | | | **0.717** | |
| **13** | **My moral ideas are much more just and fair than the ideas of those who disagree with me. †** | | | **0.756** | | | **0.703** |
| **14** | **Thoughts and behavior that conflict with my ideas about morality are wrong. †** | | | **0.656** | | | **0.583** |
| **15** | **The moral values of those who disagree with me on moral issues are probably misguided. †** | | | **0.678** | | | **0.588** |
| **16** | **Some people hold moral views that are so horrible, there is no point listening to them. †** | | | **0.406** | | | **0.353** |
| **17** | **I don't expect others to adopt the same ideas of right and wrong as me, because there will always be a diversity of moral viewpoints in the world. *** | | **0.362** | | | **0.420** | |
| **18** | **I have a better grasp on important moral topics than those who disagree with me. †** | | | **0.794** | | | **0.765** |
| **19** | **I respect that there are ways of thinking about moral issues that are different from mine. *** | | **0.588** | | | **0.627** | |
| **20** | **When another person disagrees with me on moral issues, it is highly likely that they have a mistaken view. †** | | | **0.743** | | | **0.666** |
| **21** | **I welcome different ways of thinking about important moral topics. *** | | **0.679** | | | **0.706** | |
| **22** | **My views about moral issues are just as likely to be wrong as other views. ϕ** | **0.483** | | | **0.599** | | |
| **23** | **I recognize that my views about moral issues are based on limited evidence. ϕ** | **0.668** | | | **0.713** | | |
| 24 | Although I have particular views about moral issues, I realize that I don't know everything that I need to know about it. | 0.412 | 0.343 | | | 0.422 | |
| **25** | **It is quite likely that there are gaps in my understanding about moral issues. ϕ** | **0.678** | | | **0.596** | | |

Table 2 (cont'd)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **26** | **My sources for information about moral issues might not be the best. ϕ** | **0.725** | | | **0.783** | | | |
| **27** | **I am open to new information in the area of moral issues that might change my view. *** | | **0.664** | | | **0.611** | | |
| **28** | **My views about moral issues today may someday turn out to be wrong. ϕ** | **0.633** | | | **0.721** | | | |
| **29** | **When it comes to my views about moral issues, I may be overlooking evidence. ϕ** | **0.698** | | | **0.696** | | | |
| 30 | My views about moral issues may change with additional evidence or information. | 0.401 | 0.465 | | 0.514 | 0.373 | | |
| 31 | I realize that I fall short of my own moral standards, and that there is room for improvement. | 0.549 | | | | 0.351 | | 0.553 |
| 32 | I am not always able to act in accordance with my values and principles. | 0.641 | | | | | | 0.527 |
| 33 | I consistently live up to my moral values and principles. | 0.522 | -0.49 | | | -0.34 | | 0.361 |
| 34 | My behavior and actions are consistently in line with my moral beliefs. | 0.468 | -0.487 | | 0.332 | -0.402 | | 0.302 |
| 35 | **Moral values and principles that I disagree with may be more appropriate in important moral situations. ϕ** | **0.372** | | | **0.465** | | | |
| 36 | Moral values and principles that I hold are usually appropriate for most important moral decisions. | | -0.342 | 0.356 | 0.344 | -0.427 | 0.307 | |
| **37** | **My moral values and principles may be flawed and incomplete. ϕ** | **0.734** | | | **0.640** | | | |
| **38** | **I am open to learning about different moral values from people I typically disagree with. *** | | **0.668** | | | **0.695** | | |
| **39** | **I am a more virtuous and righteous person than most others. †** | | | **0.603** | | | **0.715** | |
| **40** | **I have a stronger sense of what is right and wrong than most people. †** | | | **0.684** | | | **0.732** | |
| **41** | **I am more committed to moral values and ethical behavior than most people. †** | | | **0.713** | | | **0.751** | |
| **42** | **My moral choices and behaviors contribute more to the greater good than most people's behavior. †** | | | **0.684** | | | **0.700** | |

Table 2 (cont'd)

| | | | | | | |
|---|---|---|---|---|---|---|
| **43** | **I am not confident that my moral choices have a positive impact on the world. ϕ** | **0.339** | | **0.380** | | |
| 44 | I don't always make the right moral decisions. | 0.617 | | | | 0.571 |
| 45 | I have much to learn from other people's moral choices and behaviors. | 0.328 | 0.501 | | 0.575 | |

*Note:* Moral Fallibility (marked with ϕ signs): Items 5,7,8, 22, 23, 25, 26, 28, 29, 35, 37, 43.
Moral Learning/Openness (marked with ∗ sign): Items 9, 10, 11, 12, 17, 19, 21, 27, 38.
Moral Superiority (marked with † sign): Items 13, 14, 15, 16, 18, 20, 39, 40, 41, 42.
Factor loadings less than 0.3 are indicated by leaving space blank.

**Table 3**

*Final extracted three factors with corresponding items*

| | |
|---|---|
| | Moral Fallibility |
| 1 | My beliefs about moral issues may be incorrect. |
| 2 | I realize that my perspective about moral issues may be wrong. |
| 3 | My views about moral issues are just as likely to be wrong as other views. † |
| 4 | I recognize that my views about moral issues are based on limited evidence. |
| 5 | It is quite likely that there are gaps in my understanding about moral issues. |
| 6 | My sources for information about moral issues might not be the best. |
| 7 | When it comes to my views about moral issues, I may be overlooking evidence. †ϕ |
| 8 | My views about moral issues today may someday turn out to be wrong. †ϕ |
| 9 | Moral values and principles that I disagree with may be more appropriate in important moral situations. |
| 10 | My moral values and principles may be flawed and incomplete. |
| 11 | I am not confident that my moral choices have a positive impact on the world. |
| | Moral Openness & Learning |
| 1 | I'd be excited to learn from others about moral issues. |
| 2 | I enjoy hearing diverse perspectives about moral issues. † |
| 3 | I am interested to learn how other people think about moral issues. |
| 4 | I am open to exploring moral topics more in the future. |
| 5 | I respect that there are ways of thinking about moral issues that are different from mine. †ϕ |
| 6 | I welcome different ways of thinking about important moral topics. |
| 7 | I am open to learning about different moral values from people I typically disagree with. †ϕ |
| 8 | I am open to new information in the area of moral issues that might change my view. |
| 9 | I don't expect others to adopt the same ideas of right and wrong as me, because there will always be a diversity of moral viewpoints in the world. |
| | Moral Superiority |
| 1 | My moral ideas are much more just and fair than the ideas of those who disagree with me. †ϕ |
| 2 | Thoughts and behavior that conflict with my ideas about morality are wrong. |
| 3 | Some people hold moral views that are so horrible, there is no point listening to them. |
| 4 | The moral values of those who disagree with me on moral issues are probably misguided. |
| 5 | I have a better grasp on important moral topics than those who disagree with me. |
| 6 | When another person disagrees with me on moral issues, it is highly likely that they have a mistaken view. ϕ∗ |
| 7 | I am a more virtuous and righteous person than most others. |
| 8 | I have a stronger sense of what is right and wrong than most people. |
| 9 | I am more committed to moral values and ethical behavior than most people. |
| 10 | My moral choices and behaviors contribute more to the greater good than most people's behavior. † |

Table 3 (cont'd)
*Note*: Items included in Sample 4 are marked with † those included in Sample 6 are marked with ϕ . Items included in Sample 4 were also used in Experiment 1 and 2 in Section II, with the addition of one more item marked with ∗. An item "I understand that my moral beliefs about these statements may be limited" was additionally extracted for moral fallibility but was removed in later studies (Sample 5) after realizing that it had weird wording, resulting in a final 30-item measure. This item was thus included in survey in Sample 2, 3, and 7 but not included in any analysis.

## Scale Development: Confirmatory Factor Analysis

### Measures

#### *Moral Humility*

The full 30-item measure extracted using EFA (in Table 3) was included in four subsequent samples (Samples 2, 3, 5, 7) and Confirmatory Factor Analysis (CFA) was performed on each of them.

### Analysis and Results

The three-factor structure was replicated using Confirmatory Factor Analysis (CFA).[3] Additionally, the three-factor structure was also tested against two and one factor structure. A higher order structure was also explored.

CFA were done using the *lavaan* package for R Statistical Software (Rosseel, 2012). Structural equation modeling based on maximum likelihood estimation was used to estimate and compare the three-factor model with one-factor and two-factor models. That is, three models were tested: one factor model, two factor model, and three factor model. The one-factor model had all items loading on a single factor, a two-factor model had positively worded items (all items from moral learning/openness and moral fallibility subscales) and negatively worded items (all items from moral superiority subscale) loading on two separate factors, and a three-factor model had the items load on the factors identified in Sample 1. A hierarchical model was also tested in which all three factors loaded on a higher order factor.

The fit statistics are noted in Table 4 for different samples. Smaller relative chi square ($\chi^2$/df, relevant statistics noted in first three columns)[4], higher CFI and TLI (≥.9), smaller RMSEA and SRMR (<.09), and smaller AIC and BIC indicate better fit (Bentler 1990; Hu & Bentler, 1999; Schumacker & Lomax, 2010; Hooper et al.,

[3] Samples 4 and 6 used a short version of the moral humility scale. Because each factor did not have 3 or more items, I did not perform a CFA.
[4] Some consider significance value of the chi square test as the criteria important to judge model fit wherein non-significant p-values are considered to indicate better model fit. However, *p*-value has been found to be heavily biased by sample size, with bigger samples almost always leading to significant ($p$ <.05) values, hence not providing good information on model fit. To circumvent this problem, relative chi-square value ($\chi^2$/df ) that adjusts for sample size has been suggested as a better criterion for evaluation for model fit where lower is better.

2008). In line with this, it was observed that the three-factor model and hierarchical model were a better fit across samples compared to the two and one factor models. Consistently, chi-square difference test, used to compare the fit of two models, was also significant and indicated that the three-factor and higher order solution were better fit compared to the one and two factor solution across all studies. Table 5 has the model comparisons. The results for the hierarchical model with all three factors loading on a higher order factor showed similar fit indices as the three-factor model (the rows named "Higher" under each sample in Table 4). [5] Because the hierarchical model fits the intended conceptualization and use of moral humility measure as a general scale with subscales, this was the model of preference. A good model fit of this higher order model lends confidence to such a conceptualization and application. All items in the three-factor and higher order models had loadings of >.3 on their respective factors (see SOM).

Finally, the correlations between the three factors from CFA are in Table 6. The three factors were all significantly correlated ($p$ <.001). Moral Openness/Learning was positively correlated with Moral Fallibility ($M_r$= .66 across samples) and negatively correlated with Moral Superiority ($M_r$= -.23). Moral Fallibility was negatively correlated with Moral Superiority ($M_r$= -.41).

**Summary of Scale Development**

The exploratory factor analysis identified a thirty-item scale with three factors or subscales: moral fallibility, moral learning/openness, and moral superiority. The confirmatory factor analysis (a) replicated the three-factor structure of moral humility across different samples, (b) showed that the three-factor structure as well as the hierarchical structure with the three factors loading onto a higher order factor had decent fit, and significantly better fit than one or two factor structures, and (c) showed that the three factors are significantly correlated with each other, with moral learning/openness and fallibility correlated most, followed by moral fallibility and moral superiority, and then moral learning/openness and moral superiority. Although the three-factor and hierarchical models showed decent fit across samples, there remains room for improvement in the fit indices (e.g., >.95 CFI and TLI), something that can be explored in future work.

[5] The model showed negative variance for the moral fallibility factor in all samples, so the error variance of the factor was fixed to zero.

**Table 4**

*Robust fit indices for Samples 2,3,6, and 7*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sample 2 | | | | | | | | | | |
| | $\chi^2$ | df | *p* | $\chi^2$/df | CFI | TLI | RMSEA | SRMR | AIC | BIC |
| 1 Factor | 4748.37 | 405 | <.001 | 11.52 | 0.585 | 0.554 | 0.149 | 0.151 | 49256.50 | 49507.18 |
| 2 Factor | 2861.89 | 404 | <.001 | 7.07 | 0.765 | 0.747 | 0.112 | 0.096 | 47372.02 | 47626.87 |
| 3 Factor | 1202.43 | 402 | <.001 | 2.99 | 0.924 | 0.917 | 0.064 | 0.070 | 45716.56 | 45979.77 |
| Higher | 1203.06 | 403 | <.001 | 2.99 | 0.924 | 0.917 | 0.064 | 0.070 | 45715.19 | 45974.23 |
| Sample 3 | | | | | | | | | | |
| | $\chi^2$ | df | *p* | $\chi^2$/df | CFI | TLI | RMSEA | SRMR | AIC | BIC |
| 1 Factor | 10606.37 | 405 | <.001 | 26.19 | 0.634 | 0.607 | 0.133 | 0.133 | 140721.04 | 141036.33 |
| 2 Factor | 6308.32 | 404 | <.001 | 15.61 | 0.788 | 0.772 | 0.102 | 0.088 | 136424.99 | 136745.54 |
| 3 Factor | 2930.54 | 402 | <.001 | 7.29 | 0.909 | 0.902 | 0.067 | 0.073 | 133051.21 | 133382.27 |
| Higher | 2935.01 | 403 | <.001 | 7.28 | 0.909 | 0.902 | 0.067 | 0.073 | 133053.68 | 133379.49 |
| Sample 5 | | | | | | | | | | |
| | $\chi^2$ | df | *p* | $\chi^2$/df | CFI | TLI | RMSEA | SRMR | AIC | BIC |
| 1 Factor | 5208.35 | 405 | <.001 | 13.6 | 0.535 | 0.501 | 0.121 | 0.141 | 77301.17 | 77583.36 |
| 2 Factor | 3085.14 | 404 | <.001 | 8.2 | 0.741 | 0.721 | 0.090 | 0.098 | 75179.97 | 75466.86 |
| 3 Factor | 1522.66 | 402 | <.001 | 4.11 | 0.892 | 0.883 | 0.058 | 0.067 | 73621.48 | 73917.79 |
| Higher | 1534.19 | 403 | <.001 | 4.14 | 0.891 | 0.882 | 0.059 | 0.070 | 73631.01 | 73922.61 |
| Sample 7 | | | | | | | | | | |
| | $\chi^2$ | df | *p* | $\chi^2$/df | CFI | TLI | RMSEA | SRMR | AIC | BIC |
| 1 Factor | 6265.17 | 405 | <.001 | 15.47 | 0.608 | 0.579 | 0.135 | 0.138 | 78414.31 | 78695.16 |
| 2 Factor | 3714.35 | 404 | <.001 | 9.19 | 0.779 | 0.762 | 0.101 | 0.094 | 75865.49 | 76151.02 |
| 3 Factor | 1786.60 | 402 | <.001 | 4.44 | 0.907 | 0.900 | 0.066 | 0.072 | 73941.74 | 74236.63 |
| Higher | 1789.83 | 403 | <.001 | 4.44 | 0.907 | 0.900 | 0.066 | 0.072 | 73942.97 | 74233.18 |

*Note:* Chi square difference test was significant; three factor and higher order models were significantly better than two and one factor CFA models. See Table 5.

**Table 5**

*Model Comparison for Samples 2,3,6, and 7*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Higher vs 1-factor | | | | | |
| | $\chi^2$ Diff | df | *p* | ΔCFI | ΔTLI | ΔRMSEA | ΔSRMR | ΔAIC | ΔBIC |
| Sample 2 | 3545.3 | 2 | <.001 | -0.339 | -0.363 | 0.085 | 0.081 | 3541.306 | 3532.95 |
| Sample 3 | 7671.4 | 2 | <.001 | -0.275 | -0.295 | 0.067 | 0.061 | 7667.354 | 7656.844 |
| Sample 5 | 3674.2 | 2 | <.001 | -0.355 | -0.381 | 0.062 | 0.071 | 3670.161 | 3660.755 |
| Sample 7 | 4475.3 | 2 | <.001 | -0.299 | -0.321 | 0.069 | 0.062 | 4471.34 | 4461.978 |
| | | | | Higher vs 2-factor | | | | | |
| | $\chi^2$ Diff | df | *p* | Δ CFI | Δ TLI | Δ RMSEA | Δ SRMR | Δ AIC | Δ BIC |
| Sample 2 | 1658.8 | 1 | <.001 | -0.158 | -0.17 | 0.048 | 0.026 | 1656.822 | 1652.644 |
| Sample 3 | 3373.3 | 1 | <.001 | -0.121 | -0.13 | 0.035 | 0.015 | 3371.308 | 3366.053 |
| Sample 5 | 1551.1 | 1 | <.001 | -0.150 | -0.161 | 0.032 | 0.027 | 1548.957 | 1544.254 |
| Sample 7 | 1924.5 | 1 | <.001 | -0.129 | -0.138 | 0.036 | 0.017 | 1922.52 | 1917.84 |

*Note*: The 3-factor model had similar model comparisons with 1 and 2 factor model as the hierarchical models. Therefore, only the model comparisons with the hierarchical model is shown for space consideration.

**Table 6**

*Correlations between the three Moral Humility factors or subscales (Moral Learning/Openness, Moral Fallibility, Moral Superiority) in Samples 2,3,6, and 7*

| | | |
|---|---|---|
| | Sample 2 | |
| | Moral Openness/Learning | Moral Fallibility |
| Moral Fallibility | 0.66 | |
| Moral Superiority | -0.27 | -0.44 |
| | Sample 3 | |
| | Moral Openness/Learning | Moral Fallibility |
| Moral Fallibility | 0.73 | |
| Moral Superiority | -0.27 | -0.42 |
| | Sample 5 | |
| | Moral Openness/Learning | Moral Fallibility |
| Moral Fallibility | 0.57 | |
| Moral Superiority | -0.10 | -0.33 |
| | Sample 7 | |
| | Moral Openness/Learning | Moral Fallibility |
| Moral Fallibility | 0.68 | |
| Moral Superiority | -0.26 | -0.45 |

## Nomological Associations

Having constructed a moral humility scale and replicated the factor structure, the next aim was to explore the scale's association with other psychological constructs. Specifically, I tested if it is related to other constructs in theoretically expected and sensible ways. For instance, I expected the moral humility scale to be positively associated with constructs that capture humility (e.g., modesty, intellectual humility), open-mindedness and flexibility (e.g., openness to experience, intellectual humility, need for cognition), and positive other-orientedness (e.g., more agreeableness, less psychopathy, less narcissism). I also expected it to be negatively associated with measures that captured absolutism and extremism (e.g. political extremity, religious exclusiveness, moral grandstanding, moral absolutism).

To capture these nomological associations, personality (HEXACO, BFI), dark triad (psychopathy, Machiavellianism, narcissism), self-esteem, political identity (partisan and ideological identity), political extremity (partisan and ideological extremity), religiosity, religious exclusivism, moral grandstanding, moral relativism, attitude-specific moral conviction, intellectual humility, and need for cognition were measured. These measures were included in different samples and sometimes in more than one sample. Results are presented together in this subsection from across the samples. Example items are given below for the measures but see SOM for the full list of items wordings.

### Measures

#### *Personality, Dark Triad, and Self-Esteem*

Measures of personality were included as the expectation was that moral humility will be positively associated with the modesty/humility and openness to experience personality traits, given the overlap of different moral humility facets with these constructs. I also expected moral humility to be negatively associated with Dark Triad traits as moral humility involves positive other-orientedness, such as lack of moral superiority or moral disdain towards others. Specifically, narcissism also has associations with ego and pride (Miller et al., 2021; Tracy et al., 2009), which I especially expected to be negatively associated with moral humility. Self-esteem was included as I wanted to test that moral humility is not a lack of self-esteem, given that humility is sometimes confused with low self-regard (Tangney, 2000).

Sample 2 included the short 60-item HEXACO measure (Ashton & Lee, 2009) used to assess the six personality dimensions of honesty-humility ($\alpha = 0.80$), emotionality ($\alpha = 0.81$), extraversion ($\alpha = 0.85$),

agreeableness (α = 0.81), conscientiousness (α = 0.82), and openness to experience (α = 0.84). Sample 5 included the short 15-item Big Five Inventory (BFI-S; Soto & John, 2017) used to assess the five personality dimensions of extraversion (α = 0.63), agreeableness (α = 0.61), conscientiousness (α = 0.45), openness to experience (α = 0.61), and neuroticism (α = 0.67). Additionally, Sample 5 also included the 12-item Dark Triad scale assessing psychopathy (α = 0.76), Machiavellianism (α = 0.75), narcissism (α = 0.73) (Jonason & Webster, 2010), and 1-item self-esteem scale (Robins et al., 2001). All measures were reported on a 7-point measure (*1 = strongly disagree, 7 = strongly agree*).

***Political Identification and Political Extremity***

Measures of political identification and extremity were included because these constructs have been implicated in political animus in previous work (Brandt & Vallabha, 2024; Ganzach & Schul, 2021). I was especially interested in moral humility's relationship with political extremity given that I expected negative associations between moral humility and constructs that captured an orientation towards extremism. This is because moral humility is a construct that taps into a willingness to accept one's limitations or a willingness to be modest in one's own abilities and knowledge, and a willingness to be open to change—these are orientations that one would expect to be antithetical to extreme attitudes, positions, and values.

To assess political extremity, partisan and ideological extremity was measured. To assess political identity, partisan and ideological identity was measured. In all samples, partisan identity was measured using the 7-point partisanship scale where higher values meant stronger Republican identification (where *1 = Strong Democrat and 7 = Strong Republican*). In all samples except YouGov Sample 4, ideological identity was measured using the 7-point ideology scale where higher values meant stronger conservative identification (*1 = Very Liberal and 7 = Very Conservative*). In YouGov Sample 4, it was instead measured using a 5-point scale. In multiple samples, partisanship and/or ideology scale also had other options like "Not sure", "Don't know" and "Haven't thought much about it". In all such cases, participants who chose these options were rescored to the midpoint of the scale.

Partisan extremity was calculated by folding the partisanship scale at midpoint and forming a 4-point measure where higher value represented stronger partisan identification. Ideological extremity was calculated by folding the ideology scale at midpoint and forming a 4-point measure where higher value was strong ideological identification (except in YouGov Sample 4 where it was a 3-point measure).

***Religiosity and Religious Exclusivism/Inflexibility***

Religiosity and religious exclusivism/inflexibility was included as I expected moral humility to be negatively associated with both. I expected negative associations with religiosity as religious worldviews can be accompanied by rigidity and religion is a domain that is often linked with absolute notions of morality (Zmigrod et al., 2019; Hare, 2019; Brandt & Reyna, 2010; Hill et al., 2010). I expected negative associations with religious exclusivism given that religion often informs religious people's moral convictions — a person who considers their religion or religious views to be superior to others' might likely also consider themselves morally superior and be low in moral openness and fallibility.

In Sample 2, religiosity was measured using 6 items borrowed from Project Implicit (Schmidt et al., 2023) and another religiosity scale (Plante & Boccaccini, 1997) which were averaged to create a scale ($\alpha = 0.98$). Example items were, "I am a religious person.", "My religious faith is extremely important to me." (*1 = strongly disagree, 7 = strongly agree*).

In Sample 2, religious exclusivism (or religious inflexibility) was measured using 7 face-valid items that were borrowed from Project Implicit (Schmidt et al., 2023) or self-written. The items were averaged to create a scale ($\alpha = 0.73$). The measure intended to capture whether the participants held rigid or exclusivist religious attitudes (only one/my religion can be true). Example items were, "Everyone should have the same religious views that I have.", "There are many different religions, but only one can be true." (reverse coded). They were measured on a 7-point scale (*1 = strongly disagree, 7 = strongly agree*).

***Moral Grandstanding***

Moral grandstanding was included as it is a construct linked with morality that has been recently investigated in the context of political polarization. It has been found to be associated with more political animus, extremity, and conflicts (Grubbs et al., 2019, 2020). Moral grandstanding is understood as the use of moral speech to gain social status (Tosi & Warmke, 2020). A moral grandstander *expresses* a moral claim with the intention of being recognized by others for their moral qualities. As evident, this is a different construct from moral humility as moral grandstanding involves both status considerations and publicity consideration. Neither is part of the conceptualization of moral humility. Nevertheless, moral grandstanding may be a symptom of lower moral humility, especially considering its previous associations with both extremity and disagreeableness. Thus, I expected a negative association between moral humility and moral grandstanding.

Moral grandstanding was included in Samples 2 (α = 0.86) and 5 (α = 0.77). It was measured using 10 items on a 7-point scale (*1 = strongly disagree, 7 = strongly agree*) using a previously validated scale (Grubbs et al., 2020). Example items were "I often share my moral/political beliefs in the hope of inspiring people to be more passionate about their beliefs.", "When I share my moral/political beliefs, I do so to show people who disagree with me that I am better than them".

***Moral Relativism***

Moral relativism was included because it has also been associated with moral and political tolerance (Wright & Pölzler, 2022; Conrique, 2020; Collier-Spruel et al., 2019). I conceptualized moral humility and moral relativism to be distinct. However, these might nevertheless be psychologically congruent for some people. Indeed, previous theoretical work has suggested that very high levels of moral humility might manifest as moral relativism (Smith & Kouchaki, 2018). For these reasons, I expected moral humility to be positively associated with moral relativism.

Moral relativism was included in Samples 2 (α = 0.91), 3 (α = 0.87), 5 (α = 0.85), and 7 (α = 0.90). It was measured using 6 items on a 7-point scale (*1 = strongly disagree, 7 = strongly agree*) using a previously validated scale (Conrique, 2020). Example items were, "There is no absolute standard in morality", "What is morally good in one context may be morally bad in another".

***Moral Conviction***

Moral conviction was included as moral humility might be mistaken with a lack of moral conviction. Moral conviction, or the extent to which beliefs are moralized or treated as a matter of good/bad or right/wrong has been associated with intolerance (Skitka et al., 2021) and hence may be related to moral humility. However, moral conviction and moral humility are conceptually distinct, such that having moral humility doesn't necessarily mean not having moralized beliefs. According to my conceptualization of moral humility, a person can have moralized beliefs yet be humble about them.

Moral conviction was measured differently across samples 5, 6, and 7. Moral conviction was measured on different issues in each sample as each sample had a different main task (see the predictive validity section below) which assessed different issues for the purpose of the task. The tasks are not described in this section, only the relevant questions and measures are.

In Sample 5, participants answered their position on three randomly assigned political issues (of eleven possible issues). These issues included many topics that are moralized in politics, for example, abortion, guns, transgender rights (see SOM for full list). After answering their position on the issue, they were asked to report their moral conviction on each issue. For example, if they indicated their stance on abortion, they were asked of their moral conviction on abortion with "To what extent is your position on the issue of abortion a reflection of your core moral beliefs and convictions?" *(1= Not at All, 7 = Very Much)*. Moral conviction on the three issues were averaged to create a moral conviction measure ($\alpha = 0.78$).

In Sample 6, participants were asked to choose one of thirteen issues that was most important for them. These issues were selected on the basis of the results of a national survey run by YouGov previously in 2023 on the issues Americans found most important (Frankovic et al., 2023), such as, healthcare, immigration, etc. (see SOM for full list). After selecting their most important issue, participants were asked about their moral conviction on the issue with "To what extent is your position on [their chosen issue] a reflection of your core moral beliefs and convictions?" *(1= Not at All, 7= Very Much)*.

In Sample 7, participant's moral conviction was assessed on five issues that were the basis of the main task in that sample. These issues were guns, abortion, gender equality, racial equality, and immigration. Participants were asked, "To what extent is your position on each of the following issues a reflection of your core moral beliefs and convictions?" *(1= Not at All, 7= Very Much)*. Moral conviction on the five issues was averaged to create a measure of overall moral conviction ($\alpha = 0.79$).

***Intellectual Humility and Need for Cognition***

Intellectual humility was included as it is a type of humility that has been associated with lower levels of political polarization (Bowes, et al., 2020; Hoyle et al., 2016; Leary et al., 2017) and is also probably one of the constructs which overlaps the most with moral humility. As outlined in the introduction, I conceptualized intellectual humility and moral humility to be conceptually distinct to some extent. However, these are overlapping constructs consisting of dimensions associated with humility (e.g., openness, acknowledgement of fallibility). Thus, one can imagine a person who is generally humble in many domains, i.e., a person who is humble about their intellectual attributes and also humble about moral attributes. For these reasons, I expected moral humility to be positively associated with intellectual humility.

Intellectual humility was included in Samples 2 (α = 0.91), 3 (α = 0. 87), 5 (α = 0. 87), and 7 (α = 0.86). It was measured using 6 items on a 7-point scale (*1 = not at all true or characteristic of me, 7 = extremely true or characteristic of me*) using a previously validated scale (Leary et al., 2017). Example items were "I question my own opinions, positions, and viewpoints because they could be wrong.", "I like finding out new information that differs from what I already think is true."

Need for cognition is another construct related to people's epistemic tendencies, specifically one's enjoyment of activities that require thinking. It was included given that moral humility has aspects of open mindedness and learning, both of which have been associated with need for cognition along with lower dogmatism (Cacioppo & Petty, 1982; Furnham & Thorne, 2013; Olson et al., 1984; Liu & Nesbit, 2023). The Need for Cognition scale was used (α = 0.70; Coelho et al., 2020) in Sample 5, assessing 6 items[6] on a 7-point scale (1 = strongly disagree, 7 = strongly agree). Example items were, "I would prefer complex to simple problems.", "I like to have the responsibility of handling a situation that requires a lot of thinking."

**Analysis and Results**

Before estimating correlations of the total moral humility scale with other constructs, the items from moral superiority subscale were reverse scored and combined with items from moral fallibility and moral openness/learning subscales such that higher values on the moral humility scale indicated overall higher moral humility. Higher values on the moral fallibility subscale indicated higher moral fallibility, higher values on the moral superiority subscale indicated higher moral superiority, higher values on the moral learning/openness subscale indicated higher moral openness/learning.

For any measure that was assessed in the same way in multiple samples (e.g., partisan extremity, intellectual humility), I present a meta-analytic estimate pooled from data across samples and computed using *meta* package in R (Balduzzi, Rücker, and Schwarzer, 2019). For measures that were assessed just once or differently across sample (moral conviction, personality), estimates from the relevant samples is presented.

***Personality, Dark Triad, and Self-Esteem***

The correlations between moral humility scale and its subscales with HEXACO, BFI, Dark Triad, and self-esteem from Samples 2 and 5 are in Table 7. The correlations show that higher moral humility is significantly

[6] Two items were contributing to negative reliability for the scale due to its negative correlations with other items, so they were removed, and final analyses were done using the rest four items.

associated with higher agreeableness with mean correlation of $M_r$= 0.14, and higher openness to experience with mean correlation of $M_r$ = 0.22, across HEXACO and BFI. These correlations were primarily driven by moral learning/openness subscale and then moral fallibility subscales, i.e., agreeableness and openness to experience were most strongly and reliably correlated with the moral learning/openness part of the moral humility, followed by moral fallibility.

Moral humility scale was not correlated with the honesty-humility subscale of the HEXACO inventory. However, the modesty facet of the honesty-humility subscale, which captures the humility aspect and was of my interest was positively associated with moral humility ($r$ = 0.16, $p$ <.001). This was primarily driven by a negative correlation of modesty with moral superiority subscale ($r$ = - 0.36, $p$ <.001), suggesting that humility/modesty captured in HEXACO is most closely related to the moral superiority subscale. This makes sense as the items capturing modesty (or humility) assessed how much people think they are better than others (e.g., "I think that I am entitled to more respect than the average person is") which is most similar to the content of the moral superiority subscale.

Correlations between moral humility scale and the Dark Triad (psychopathy, Machiavellianism, narcissism) revealed that higher moral humility was significantly negatively associated with all three dark traits, psychopathy, Machiavellianism, and narcissism, with similar pattern of correlations at subscale level too. However, there was an unexpected small positive correlation between the moral fallibility subscale and psychopathy. This might indicate that people with higher scores on moral fallibility are also more willing to report that they have moral failings (such as the traits included in psychopathy). Finally, the moral humility scale was not significantly associated with self-esteem.

In summary, the correlations indicate that the moral humility scale taps into the constructs of modesty/humility, openness, and other-orientedness (higher agreeableness and lower levels of dark traits) as was expected. Moral humility did not tap into self-esteem, indicating humility is not simply self-deprecation or a lack of self-regard (Tangney, 2000).

**Table 7**

*Correlations between the Moral Humility scale and subscales with HEXACO, Big 5, Dark Triad, and Self Esteem scales*

| | Overall | Subscales | | |
|---|---|---|---|---|
| | HEXACO (Sample 2) | | | |
| | Moral Humility | Moral Learning/ Openness | Moral Fallibility | Moral Superiority |
| Modesty | 0.16*** | 0.03 | 0.01 | -0.36*** |
| Honesty-Humility | 0.01 | 0.01 | -0.09* | -0.13** |
| Emotionality | 0.08 | 0.05 | 0.09 | -0.03 |
| Extraversion | -0.04 | 0.11* | -0.11* | 0.08 |
| Agreeableness | 0.15*** | 0.24*** | 0.08 | -0.05 |
| Conscientiousness | -0.06 | 0.11* | -0.17*** | 0.03 |
| Openness to Experience | 0.27*** | 0.40*** | 0.19*** | -0.07 |
| | BFI (Sample 5) | | | |
| Extraversion | -0.07* | 0.05 | -0.03 | 0.17*** |
| Agreeableness | 0.13*** | 0.34*** | <.01 | 0.04 |
| Neuroticism | 0.02 | 0.06 | 0.05 | 0.05 |
| Conscientiousness | -0.03 | -0.05 | 0.10** | 0.12*** |
| Openness to Experience | 0.17*** | 0.42*** | 0.07* | 0.11** |
| | Dark Triad (Sample 5) | | | |
| Machiavellianism | -0.11*** | -0.15*** | 0.06 | 0.17*** |
| Psychopathy | -0.12*** | -0.25*** | 0.09** | 0.12*** |
| Narcissism | -0.13*** | -0.06 | 0.04 | 0.28*** |
| | Self-Esteem (Sample 5) | | | |
| Self Esteem | -0.04 | 0.03 | -0.04 | 0.06 |

*Note:* $^{*}p<0.05$; $^{**}p<0.01$, $^{***}p<0.001$

***Political Identification and Political Extremity***

Meta-analytic correlations of moral humility with partisan identity (higher scores indicated stronger Republican identity), ideological identity (higher scores indicated higher conservatism), partisan extremity, and ideological extremity (extremity is distance from the middle of the scale) are in Table 8. The mean correlations across samples show that moral humility is significantly and negatively associated with conservatism ($M_r$= - 0.25), Republican identity ($M_r$= - 0.18), partisan extremity ($M_r$ = -0.12), and ideological extremity ($M_r$= -0.19). The subscales were largely consistent with the results for entire scale. In summary, the correlations indicate that the moral humility scale taps into lower extremist inclinations, as was expected. There was also a positive relationship between moral humility and liberal ideological leanings (alternatively negative relationship between moral humility

and conservatism) which although not predicted a priori, might be moral humility tapping into openness given that liberal ideology has previously been linked with openness to experience (Sibley et al., 2012)

**Table 8**

*Meta-analytic correlations between the Moral Humility scale and subscales with Ideological Extremity, Partisan Extremity, Ideological Identity, Partisan Identity, Moral Relativism, Intellectual Humility, and Moral Grandstanding*

| | | | Overall | Subscales | | |
|---|---|---|---|---|---|---|
| | k | N | Moral Humility | Moral Learning/ Openness | Moral Fallibility | Moral Superiority |
| Ideological Extremity | 6 | 5670 | -0.19** | -0.11** | -0.18** | 0.16** |
| Partisan Extremity | 6 | 5647 | -0.12** | -0.06** | -0.10** | 0.11** |
| Ideological Identity (higher conservatism) | 6 | 5559 | -0.25** | -0.24** | -0.23** | 0.08 |
| Partisan Identity (stronger Republican) | 6 | 5646 | -0.18** | -0.18** | -0.17** | -0.07 |
| Moral Relativism | 4 | 3661 | 0.61** | 0.57** | 0.57** | -0.27** |
| Intellectual Humility | 4 | 3670 | 0.56** | 0.65** | 0.47** | -0.17** |
| Moral Grandstanding | 2 | 1392 | -0.27** | -0.04** | -0.09** | 0.47** |

*Note:* $^{*}p<0.05$; $^{**}p<0.01$, $^{***}p<0.001$

***Religiosity and Religious Exclusivism/Inflexibility***

Correlations of moral humility with religiosity and religious exclusivism from Sample 2 are presented in Table 9. The correlations show that moral humility is significantly and negatively associated with religiosity and religious exclusivism. The subscales were consistent with the results for entire scale. The correlations with religiosity and religious exclusivism indicate that the moral humility scale taps into lower rigidity, superiority, and absolutism.

***Moral Grandstanding***

Meta-analytic correlations of moral humility with moral grandstanding are in Table 8. Moral humility was significantly and negatively associated with moral grandstanding ($M_r$ = -0.27) with the strongest associations with moral superiority subscale. In summary, the correlations with moral grandstanding indicate that the moral humility scale taps into lower levels of status-seeking, sanctimoniousness, extremity, and superiority.

***Moral Relativism***

Meta-analytic correlations of moral humility with moral relativism are in Table 8. Moral humility was significantly and positively associated with moral relativism ($M_r$ = 0.61). The subscales were largely consistent with the results for entire scale, with moral openness/learning and moral fallibility subscales showing the strongest

correlations. In summary, the correlations with moral relativism indicate that the moral humility scale taps into lower absolutism. It also suggests that moral relativism and moral humility though conceptually distinct are empirically the most overlapping constructs.

***Moral Conviction***

Correlations of moral humility with moral conviction from Samples 5, 6, and 7 are in Table 9 (not pooled because they were measured differently). The correlations show that moral humility is not significantly associated with moral conviction. However, there were small significant positive correlations with the moral learning/openness subscale and the moral superiority scale. In summary, the correlations with moral humility indicate that moral humility is not simply a lack of moral convictions.

**Table 9**

*Correlations between the Moral Humility scale and subscales with Religiosity, Religious Exclusivism Moral Conviction, and Need for Cognition*

| | Overall | Subscales | | |
|---|---|---|---|---|
| | Moral Humility | Moral Learning/ Openness | Moral Fallibility | Moral Superiority |
| | | Religiosity | | |
| Sample 2 | -0.23*** | -0.10* | -0.21*** | 0.23*** |
| | | Religious Exclusivism | | |
| Sample 2 | -0.47*** | -0.30*** | -0.36*** | 0.46*** |
| | | Moral Conviction | | |
| Sample 5 | -0.03 | 0.10** | -0.04 | 0.12** |
| Sample 6 | -0.01 | 0.09** | 0.01 | 0.13** |
| Sample 7 | 0.06 | 0.15** | 0.05 | 0.04 |
| | | Need for Cognition | | |
| Sample 5 | 0.16*** | 0.29*** | 0.11*** | -0.05 |

*Note:* $^{*}p<0.05$; $^{**}p<0.01$, $^{***}p<0.001$

***Intellectual Humility and Need for Cognition***

Meta-analytic correlations of moral humility with intellectual humility and need for cognition are in Table 8. Moral humility was significantly positively associated with intellectual humility ($M_r = 0.56$) and need for cognition ($r = 0.16$). The subscales were largely consistent with the results for entire scale, with the moral openness/learning subscale showing the strongest correlations across samples. In summary, the correlations with intellectual humility and need for cognition indicate that the moral humility scale taps into more openness, flexibility, acceptance of fallibility, and willingness to learn.

**Summary of Nomological Associations**

The correlations suggest that moral humility was associated with constructs in sensible and largely expected ways. Taken together, the correlations indicate that a person high in moral humility is inclined towards less extremism and absolutism (e.g., less political extremity, religious exclusiveness, moral grandstanding, moral absolutism, religiosity), shows openness and flexibility (e.g., openness to experience, intellectual humility, need for cognition), shows humility (e.g., modesty, intellectual humility), and positive orientation towards others (e.g., more agreeableness, lower psychopathy, lower narcissism, etc.). They also suggest that moral humility is not simply a lack of self-esteem or moral conviction.

The positive correlations with modesty and intellectual humility provided evidence for moral humility's convergent validity. The lack of correlations with self-esteem and small correlations with moral conviction provide evidence of moral humility's discriminant validity. Together, the correlations provide important empirical evidence that the moral humility scale is working as it should and provide insight into the nature of the moral humility construct.

The strength of the various nomological associations varied, with moral humility having the strongest correlations with moral relativism and intellectual humility ($r \sim 0.5$-$0.6$). This is likely due to both having the greatest content and/or construct overlap with moral humility. This suggests that these constructs are very close to moral humility in the nomological network, and it is important for future studies to establish the incremental validity of these constructs when explaining relevant outcomes. This could help clarify whether these constructs are distinct enough to be meaningful. I examine this further in the following sections.

The correlations with other constructs, such as modesty, agreeableness, openness to experience, dark traits, political extremity, religiosity were small to medium (~ 0.1-0.3). These smaller correlations, compared to moral relativism and intellectual humility, is likely because these other constructs share some conceptual/construct similarities with moral humility, but at a more distal level. It also suggests that moral humility is not reducible to these personality and personality-type constructs.

Interestingly, there were also small, unexpected associations of the moral fallibility subscale with psychopathy, $r = .09$. In a similar vein, although not described in main text, small negative correlations of moral fallibility with sincerity and fairness facets of honesty-humility subscale were also observed, reflected in a negative correlation of moral fallibility with honesty-humility ($r = -.09$) in Table 7. What might these small correlations

convey? All three of these constructs (psychopathy, sincerity, and fairness) captured inclination towards things that might be considered morally questionable. For example, psychopathy had participants indicate how much they agreed with, "I tend to be insensitive" sincerity had "I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed." (reverse coded), and fairness had "I'd be tempted to use counterfeit money, if I were sure I could get away with it.". A small positive correlation of moral fallibility with these suggest that moral fallibility might also be capturing some people who show inclination towards morally questionable behaviors and also *acknowledge* such inclinations as moral limitations. These correlations also suggest indirectly that moral humility scale is not capturing social desirability.

**Predicting Polarization Outcomes**

Having constructed a scale, confirmed its factor structure, and tested if it was associated with constructs in sensible and expected ways, the next aim was to investigate the central question in the project. That is, examining whether moral humility could counteract the dark features of morality as they manifest in the context of political polarization in USA. I examined if moral humility was associated with lower levels of polarization outcomes *and* if moral humility was associated with lower levels of polarization over and above the effects of other close constructs. Thus, across Samples 3-7, I tested moral humility's predictive or criteria validity and incremental validity. The negative associations between moral humility and political extremity observed thus far foreshadow that moral humility might have a counteractive effect on polarization. In the next steps, I investigated this link across additional indicators of polarization.

In Sample 3, the aim was to test if moral humility predicted lower polarization (e.g., animus towards political outgroup, perceptions of threat from political outgroup, anger and negative affect towards political outgroup, social distance from political outgroup) as well as anti-democratic outcomes, especially when misperceptions of outpartisans are corrected. Previous research has found that people often hold inaccurate perceptions of outpartisans. These misperceptions can be about the characteristics (e.g., demographic) of outpartisans (Ahler & Sood, 2018), about how hostile, prejudiced, or negative outpartisans are towards one's ingroup (Moore-Berg & Hameiri, 2020; Ruggeri et al., 2021), or about outpartisans' policy positions, values, or support for ethically questionable actions (Enders & Armaly, 2019; Pasek et al., 2022; Mernyk et al., 2022; Moore-Berg & Hameiri, 2024). For example, people hold inaccurate perceptions about the extent to which political outgroup would engage in violence or anti-democratic actions (Pasek et al., 2022; Mernyk et al., 2022). Such

misperceptions have been suggested to drive hostility towards political outgroups (Moore-Berg & Hameiri, 2024). This is evidenced by work suggesting that correcting such misperceptions reduces negativity towards political outgroups (Moore-Berg & Hameiri, 2024; Voelkel et al., 2023).

Adapting a misperception correction task used in previous work (Lees & Cikara, 2020; Voelkel et al., 2023), I investigated if higher moral humility would predict lower polarization and anti-democratic attitudes, and if this would be greater when people were given a misperception correction treatment (compared to control). The misperceptions corrected were about hostility of the outgroup towards the ingroup. Apart from polarization outcomes, I also tested anti-democratic attitudes as outcomes because polarization has been linked with a decline in support for democratic practices (Finkel et al., 2020). Notably, utilizing a misperception correction task also allowed to test how moral humility is linked with people's misperceptions about political outgroups. In sum, this study tested the relationship of moral humility with polarization, anti-democratic attitudes, and misperceptions. Sample 4 which was a national YouGov sample had pre-included measures of these three outcomes which allowed conceptual replication of these three relationships using slightly different measures and design (e.g., misperceptions were about outgroup's democratic commitments).

In Sample 4 and 5, the primary aim was to test if higher moral humility predicted more openness towards opposing political standpoints. Previous work has found that all political sides engage in avoidance of information or opinions from the other side (Frimer et al., 2017), and that lack of cross-cutting information amplifies polarization (Hobolt et al., 2023; Levy, 2021). Adapting a selective exposure task used in previous work (Frimer et al., 2017), I investigated if higher moral humility would predict people being more interested in learning about opposing political viewpoints from people who disagreed with them. The political issues included in the studies included multiple topics frequently moralized in politics (e.g., voting for Trump/Biden, abortion, guns, transgender, healthcare, etc). A secondary aim in Sample 4 was to conceptually replicate the relationship of moral humility with polarization, anti-democratic attitudes, and misperceptions as aforementioned.

In Sample 6, the aim was to test if higher moral humility predicted more support for political compromise. Compromise serves as a means to recognize, respect, and accommodate pluralistic values in a democracy. Previous work has found that polarized and moralized individuals reject political compromise (Clifford, 2019; Delton et al., 2020; Kodapanakkal et al., 2022; Ryan, 2017; Finkel et al., 2020). Thus, adapting a political compromise task used

in previous work (Kodapanakkal et al., 2022), I investigated if higher moral humility facilitates more favorable attitudes towards political compromise.

In Sample 7, the aim was to test if higher moral humility predicted lesser sharing of partisan-consistent news online (or myside sharing) on moralized topics like abortion, gun, immigration, race, and gender. Previous work has found that moralized and politically extreme individuals are more likely to share partisan-consistent news online (Marie et al., 2023). Research also shows that social media aggravates the negative facets of morality such as moral outrage (Van Bavel et al., 2024). Thus, adapting a myside sharing task used in previous work (Marie et al., 2023),I investigated if higher moral humility tempers inclination to share partisan and moralized news on social media.

**Sample Size Justification**

In Sample 3, sample size was based on estimates of misperception treatment effect and moral humility's effect on outcomes from previous work (Voelkel et al., 2023; Vallabha et al., 2024). A sample of 1500 was chosen based on power analysis using InteractionPoweR (Finsaas & Barangeras, 2018) which indicated between ~80% to ~90% power to detect a moral humility effect in the range $r$ [0.08, 0.2] after accounting for the treatment and interaction effects of $r = 0.07$ (Voelkel et al., 2023).

In Sample 4, sample size justification was based on two factors. First, the YouGov national survey where I proposed to field my study allowed a sample of 1000. Second, in Sample 3 the correlation between moral humility and polarization outcomes ranged from 0.16 to 0.27. Further, moral humility was correlated with misperceptions at $r = 0.1$ and with support for anti-democratic candidates at $r = 0.25$. Based on these prior results, G-power (Faul et al., 2007, 2009) computed that the smallest effect size observed previously of ~ 0.1 can be detected with 90% power (alpha = .05) with a sample of ~1000. This gave confidence in the ability of this study to detect the effects of interest and was preregistered.

Sample 5 was a replication of Sample 4, thus the same sample size as Sample 4 ($N = 1000$) was chosen for the study. Sensitivity analysis in G-power function indicated that the study had 80% power to detect an effect size of $r = 0.09$. A post-hoc analysis of achieved power using simr package in R (Green & MacLeod, 2015) showed that there was 100% power to detect the effects between moral humility and selective exposure.

Sample 6 had similar sample constraints as Sample 4 as it was also fielded in a YouGov national sample which only allowed for a sample of 1000. Sensitivity analysis function indicated that the study had 80% power to

detect an effect size of $r = 0.09$. A post-hoc analysis of achieved power using, G-power (Faul et al., 2007, 2009) showed that we had ~99 to 100 % power to detect the effects between moral humility and political compromise across the outcomes.

In Sample 7, power analysis using Summary Statistics Based R for multilevel analysis (Murayama et al., 2022) was conducted based on estimates from previous work (Marie et al., 2023) using the same design. Power analysis suggested that a sample of ~800 had 90% power to detect a small cross-level interaction effect size of $b = 0.07$ ($t = 2.07$).

**Measures**

***Moral Humility***

In Samples 3, 5, and 7, moral humility was measured using the full 30-item measure of moral humility (Sample 3 $\alpha = 0.94$, Sample 5 $\alpha = 0.89$, Sample 7 $\alpha = 0.94$). In Samples 4 and 6, due to the limited number of items that could be included or proposed in the respective YouGov national surveys, shortened measures were proposed. The items for the shortened measure were sampled almost equally from all subscales (Sample 4: 8-item measure, $\alpha = 0.75$; Sample 5: 6-item measure, $\alpha = 0.67$). These items were picked using genetic algorithm for creating shortened scales (Scrucca & Sahdra, 2016). The algorithm iteratively evaluates different combinations of items on various parameters (like reliability, content coverage) and eventually selects items that maximize these parameters. See Table 3 for the selected items marked using symbols.

Subscales were also used to explore relationships with outcomes at the subscale level, i.e., with Moral Learning/Openness (Sample 3 $\alpha = 0.93$, Sample 4 $\alpha = 0.75$, Sample 5 $\alpha = 0.91$, Sample 6 $\alpha = 0.71$, Sample 7 $\alpha = 0.93$), Moral Fallibility (Sample 3 $\alpha = 0.93$, Sample 4 $\alpha = 0.75$, Sample 5 $\alpha = 0.86$, Sample 6 $\alpha = 0.68$, Sample 7 $\alpha = 0.93$), Moral Superiority (Sample 3 $\alpha = 0.88$, Sample 4 $\alpha = 0.66$, Sample 5 $\alpha = 0.84$, Sample 6 $\alpha = 0.6$, Sample 7 $\alpha = 0.89$).

***Polarization or Social Cohesion Measures***

Several polarization outcomes were measured, drawing from previous work capturing intergroup perceptions and behaviors in the political context (Iyengar et al., 2019; Kuhne & Kamin, n.d; Vallabha et al., 2024). These included Social Cohesion Impact Measures (SCIM), a group of measures collated by academics and practitioners to measure depolarization (Kuhne & Kamin, n.d). Example of measures and items are in Table 10.

**Table 10**

*Example of polarization measures used in Sample 3 (including SCIM measures).*

| Construct | Example Items |
|---|---|
| Affective Polarization/ Partisan Affect Gap | How would you rate Democrats?<br>How would you rate Republicans?<br>(0= Very cold or unfavorable feeling, 100=Very warm or favorable feeling) |
| Social Distance | How comfortable are you having friends who are [outgroup members]? 0= Not at all, 10 = Extremely |
| Humanization | How often do you think [outgroup members] experience the following emotions? 0 = Never to 10 = Very frequently<br>(i) Hope.<br>(ii) Admiration. |
| Morality | Would you say that [outgroups members] are generally good people? 0= Not at all, 10 =Absolutely |
| Intergroup Empathy | I find it difficult to see things from [outgroup members] point of view. 0 = Strongly disagree, 10 = Strongly agree<br>It is important to understand [outgroup members] by imagining how things look from their perspective. 0 = Strongly disagree, 10 = Strongly agree |
| Respect/ Understanding | Even if I don't agree with them, I understand people have good reasons for voting for [outgroup] candidates. 0 = Strongly disagree, 10 = Strongly agree<br>I respect [outgroup members'] opinions even when I do not agree.<br>0 = Strongly disagree, 10 = Strongly agree |
| Pluralist Norms | How important to you is it that [ingroup] elected officials make compromises with [outgroup] elected officials to solve important problems? 0= Not at all, 10 = Extremely<br>How likely would you be to vote for a [ingroup] candidate who said they would ban [extreme outgroup] group rallies on the state capitol grounds? 0= Not at all, 10 = Extremely |
| Perceived Threat | Would you say [outgroup members] are a serious threat to the United States? 0= Not at all, 10 =Absolutely |
| Anger | How angry do you get just thinking about [outgroup members]? 0= Not at all, 10 = Extremely |
| Identity | How much do you agree with the statements "if I met someone who is a [member of ingroup], I'd feel connected to that person"?<br>0 = Strongly disagree, 10 = Strongly agree |

*Note:* Inter-item correlation for Humanization was 0.8 ($p$ <.001), Empathy was 0.5 ($p$ <.001), Respect/Understanding 0.7 ($p$ <.001), and Pluralistic Norms was 0.2 ($p$ <.001).

To give a brief overview, Sample 3 included measures capturing negative orientation towards outgroup like partisan affect gap (popularly used as a measure of affective polarization; Iyengar et al., 2019), negative affect

towards outgroup, anger towards outgroup, social distance from outgroup, perceptions of outgroup threat, and close identification with ingroup. Measures capturing positive orientation towards outgroup included perceptions of outgroup morality, empathy towards political outgroup, humanization of political outgroup, respect/understanding towards political outgroup, and support for pluralistic norms. Sample 4 had the partisan affect gap or affective polarization measure only, captured using thermometer ratings towards Republicans/Democrats.

***Anti-Democratic Attitudes***

Anti-democratic attitudes were measured in Samples 3 and 4. Sample 3 used 6 items from previous work on polarization and democratic outcomes (Voelkel et al., 2023), which were averaged together to create a measure of anti-democratic attitude ($\alpha = 0.91$). These items captured support towards political inparty candidates who engage in anti-democratic practices such as gerrymandering, disputing election legitimacy, and suppressing media. An example item was, "How likely would you be to vote for the [Inparty] candidate if you learned that they said that [Inparty] should not accept election results if they do not win?".

In Sample 4, democratic attitudes were similarly measured by taking an average of 4 items that were included in the national survey which captured support for anti-democratic norms ($\alpha = 0.74$). Participants rated how much they agreed or disagreed with various anti-democratic actions like censorship, subverting court decision or Congress/Executive, and interfering with polling. An example item was, "Do you agree or disagree with the following: The government should be able to censor media sources that spend more time attacking [inparty] than [outparty]." See SOM for full details on items from both samples.

***Misperceptions***

Misperceptions were measured *and* corrected in Sample 3, and only measured in Sample 4. In Sample 3, the misperception correction task was taken from previous work (Lees & Cikara, 2020; Voelkel et al., 2023). Therein, participants were randomly assigned to either the misperception correction treatment ($N = 732$) or control ($N = 739$) condition. In the misperception correction treatment condition, participants were randomly presented with one of five scenarios wherein their ingroup was undertaking an action that would possibly disadvantage the outgroup. An example scenario was "A [Inparty] controlled state legislature is considering a law that would require sitting governors to disclose their tax returns and all possible financial conflicts of interests. The law would go into effect immediately, and the current sitting governor is a [Outparty Person]." See SOM for all five scenarios. Participants were asked to indicate the extent to which they thought a partisan outgroup member would dislike the action, oppose

the action, and find the action politically unacceptable. These, when subtracted from outgroup members' actual dislike, opposition, and unacceptability, indexed participant's level of misperceptions[7].

The misperception correction took place next wherein the participants in the treatment condition were informed of the real responses from partisan outgroup members on how much they actually disliked/opposed/found unacceptable the actions. The real responses presented were based on a previous nationally representative survey (Lees & Cikara, 2020). The control condition did not take part in the misperception task and directly moved to answering the outcome measures.

In Sample 4, misperceptions about support for anti-democratic actions were only measured. Misperceptions about the outparty were calculated in two steps. First the percentage of Democratic and Republican participants agreeing to four anti-democratic actions described above (censorship, subverting court decision or Congress/Executive, and interfering with polling) was estimated. For example, "Do you agree or disagree with the following: The government should be able to censor media sources that spend more time attacking [inparty] than [outparty]." See SOM for full details on all items. This indexed actual support towards anti-democratic actions by Democrats and Republican respectively. In the second step, these actual agreement values for the outgroup were subtracted from each participant's response to four questions assessing participants' *perceptions* of outgroup members' approval of anti-democratic actions noted above. For example, for media censorship, they were asked, "What percent of [outparty] voters do you think agree with the following: The government should be able to censor media sources that spend more time attacking [outparty] than [inparty]." The values obtained from subtraction of actual values estimated in first step from perceived values in second step for each of the four democratic action was averaged to create measure of misperceptions ($\alpha = 0.85$). These gave estimates of the magnitude of each participant's misperception of outgroups.

***Selective Exposure***

Selective exposure, the primary outcome in Sample 4 and 5, was measured by assessing participants' interest in opposing political standpoints. For this, I adapted a selective exposure task from previous work (Study 2-3; Frimer et al., 2017). In both Samples 4 and 5, for one (Sample 4) or three (Sample 5) political issues answered by

[7] Misperceptions could only be calculated for the treatment group who actually reported their (mis)perceptions of how much they thought outgroup members would oppose, dislike, and find unacceptable certain actions. Following the materials from Voelkel and colleagues (2023), the control group did not complete these initial judgments.

participants previously in the study, participants were asked how interested they would be in hearing from someone with the opposing viewpoint on the issue.

The issues were chosen slightly differently in both samples. In Sample 4, there were ten issues that a participant could answer plus one issue that everyone had to answer. Of the ten possible issues, each participant was randomly assigned to answer only five issues. From these five issues and one mandatory issue[8] answered by everyone, only one issue was then randomly selected for selective exposure task. In Sample 5, of the eleven total issues in Sample 1, each participant answered three randomly selected issues and then completed selective exposure task for the three issues. The issues included environment, healthcare, transgender, abortion, defunding police, guns, trade, taxes, unions, marijuana, presidential vote. These issues were already part of the national survey wherein I fielded moral humility and selective exposure measures, so I leveraged participants' response to these issues for the selective exposure task. See SOM for full wording of how issues were assessed.

For example, if a participant answered the transgender issue, they were first asked, "Some believe that transgender athletes should be allowed to compete on teams that match the gender they identify with. Others believe that transgender athletes should be required to compete on teams that match the sex they were assigned at birth. Still others fall somewhere between these two positions. Where do you stand on this issue?" They then indicated their response on a 7-point scale (*1= Allow transgender athletes to compete on teams matching their gender identity, 4 = Middle of the road; see the pros and cons of both sides, 7 =Require transgender athletes to compete on teams matching their sex assigned at birth*).

If someone answered 1, 2, or 3 on the scale, they then were asked later, "You indicated that you leaned towards allowing transgender athletes to compete on teams matching their gender identity. How interested are you in hearing from someone who supports requiring transgender athletes to compete on teams matching their sex assigned at birth?". If they instead answered 5, 6, or 7, they read, "You indicated that you leaned towards requiring transgender athletes to compete on teams matching their sex assigned at birth. How interested are you in hearing from someone who supports allowing transgender athletes to compete on teams matching their gender identity?". If they answered 4, another issue was randomly selected for selective exposure task in Sample 4 wherein participant's

[8] The mandatory question was about their vote choice in previous election, which was asked of everyone at the beginning of the study

response was not a 4. In Sample 5, if participant answered 4 on any issue, they did not get the selective exposure task for that issue.

Participants then indicated their response on a -100 to 100 scale *[−100 (very uninterested), 0 (neutral), and 100 (very interested)*]. Higher values indicated more interest in learning from someone who disagreed with them politically, or more interest in cross-cutting exposure.

***Political Compromise***

The measure for political compromise, the main outcome in Sample 6, was adapted from Kodapanakkal et al. (2022). Participants were first asked to choose one of thirteen issues that was most important for them. The issues included in this list were selected on the basis of results of a national survey run by YouGov in 2023 on the issues Americans found most important (Frankovic et al., 2023), such as, abortion, healthcare, immigration, guns. (see SOM for full list). After selecting their most important issue, participants completed a political compromise task from Kodapanakkal et al. (2022) where they were asked about two candidates with different approaches to the issue they had selected as most important,

> "We would like your opinion on two candidates with different approaches to [Most Important Issue]. The candidates might be competing for their party's nomination to run for Congress. Both candidates agree with your position on [Most Important Issue], but they differ on how they plan to negotiate with their political opponents.
>
> Candidate A is uncompromising and will vote against any proposal that does not support your position.
> Candidate B will dislike proposals that do not support your position, but will be willing to negotiate and make concessions in this area if it leads to a gain in other areas that are important to you."

Participants indicated their likelihood of supporting each candidate separately, "How likely are you to support Candidate [A/B]?" using a 7-point Likert scale (*1 = not at all, 7 = very likely)*. Support for Candidate A (the uncompromising candidate), support for Candidate B (the compromising candidate), and the relative support of compromising candidate over uncompromising candidate (Candidate B - Candidate B) formed the three main outcome measures for the study.

***Myside Sharing***

The materials and measures for myside sharing or partisan sharing, the main outcome in Sample 7, was taken from Marie et al. (2023; Experiment 1 &10). Participants were randomly assigned to one of two news sets which contained 12 real news items each. Thus, in total there were 24 news items but to prevent participant fatigue, participants were assigned to only one set of news stories.

The news items were on divisive issues: gun control, abortion, gender, race, and immigration. Each set contained 12 news, two news items per these five issues, one which was congruent for liberals (e.g. pro-abortion) and one congruent for conservatives (e.g. anti-abortion). This resulted in 10 news items; additionally, there were 2 neutral news items in each set. All news items were real and had been picked from mainstream news media websites. Each news item had a headline, a short introductory snippet, and a picture. All news items are in SOM.

After reading the 12 news items, participants were asked about their inclination to share those news stories on social media, "How likely would you be to share this news article on social media?" (*1=Extremely Unlikely, 2= Unlikely, 3=Somewhat Unlikely, 4 = Somewhat Likely, 5 = Likely, 6 = Extremely Likely*).

Participants also indicated their position on the five issues on which the news items were based. For race, they were asked "What is your position on the issue of racial equality? *(0= I don't care at all, 100 = Extremely in favor)*, for gender "What is your position on the issue of gender equality?" *(0= I don't care at all, 100 = Extremely in favor)*, for abortion "What is your position on the issue of abortion?" *(0= Extremely pro-life, 100 = Extremely pro-choice)*, for guns "What is your position on the issue of guns?" *(0= Extremely pro-gun rights, 100 = Extremely pro-gun control)*, and for immigration "What is your position on the issue of immigration?" *(0=Extremely in favor, 100= Extremely opposed)*. Participants' position on the issues was used to determine which news items (liberal or conservative for each issue) was politically congruent and incongruent to participants.

The study was opened to only those participants on Prolific who indicated on their prescreening survey that they used at least one of the following social media: Facebook, Twitter, Instagram, Tik Tok, and Reddit. This choice was made based on past work on online news sharing that used similar eligibility criterion (Mosleh et al., 2020). This was to ensure that our study was realistic to participants doing the study as the study pertains to online news sharing.

***Controls***

Partisan identity, ideological identity, partisan extremity, ideological extremity, and demographics (age, race, gender) were included as the standard set of control variables in all analyses conducted in Samples 3-7. Gender (male = 0.5, female = -0.5) and ethnicity (0.5 = white, -0.5 = non-white) were contrast coded.

Samples that were not collected as part of national surveys (wherein I was limited in the number of items I could include, i.e., Sample 4 and 6) included moral relativism and intellectual humility as control variables as well to establish moral humility's incremental validity over these closely related psychological constructs. These two

were chosen as these two constructs are most closely conceptually and/or empirically related to moral humility. They were thus included in Samples 3, 5, and 7. Sample 6 included moral conviction as well as it had been preregistered as control for exploratory purposes.

**Analysis and Results**

***Sample 3 & 4: Misperceptions, Polarization, Anti-Democratic Attitudes***

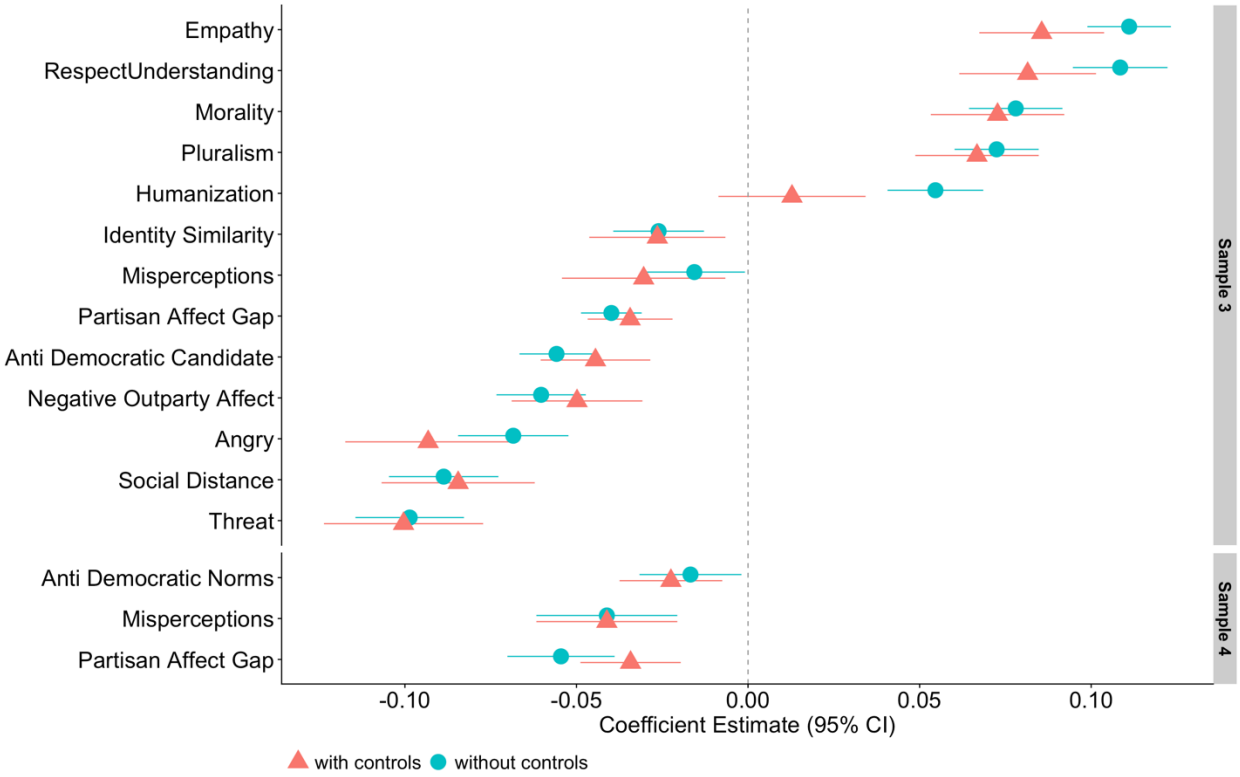
**Sample 3 Results.** In Sample 3, the social cohesion outcomes and democratic outcomes were regressed on moral humility, the condition variable (contrast coded: Control= -5, Misperception Correction Treatment = .5), and the interaction term between them for the first set of models. In the second set, the same models were re-estimated including all the covariates. I expected main effects for moral humility and the condition variable, as well as a significant interaction term between them. Moral humility was mean-centered before analyses to improve interpretability for the interaction analyses.

Unlike previous work (Lees & Cikara, 2020; Voelkel et al., 2023), the misperception correction treatment did not work as expected and didn't have a significant effect on any of the outcome measures (all $p$'s > .05 $p$ range [0.16, 0.96]). There also wasn't a significant interaction of the treatment with moral humility (all $p$'s > .05 $p$ range [0.16, 0.86]). This suggests that misperception corrections did not reduce animosity towards outgroup, and this non-significant effect was similar across levels of moral humility. There were however main effects of moral humility on the polarization and democratic outcomes across conditions. Additionally, moral humility was also associated with lower magnitude of misperceptions for the experimental group (the control group did not complete these measures). These associations are shown in Figure 2 (upper part).

Higher moral humility was associated with lower partisan affect gap (or affective polarization), lower desire to socially distance from the political outgroup, lower negative affect and anger towards the political outgroup, lower perceptions of threat from the political outgroup, lower misperceptions of the outgroup members, and lower support towards anti-democratic candidates. Higher moral humility was also associated with more empathy, humanization, and respect/understanding of political outgroup, more support for pluralistic norms, and higher perceptions of outgroup morality. These relationships were robust to the inclusion of control variables like intellectual humility and moral relativism (except for humanization).

**Figure 2**

*Sample 3 and 4 Relationship between Moral Humility and Political Outcomes*




*Note:* The blue lines with circles indicate models where moral humility was the only predictor; the orange lines with triangles indicate the models where all controls were also included. Outcomes were recoded to range from 0-1. The controls in Sample 3 included the standard controls plus intellectual humility and moral relativism, controls in Sample 4 only included the standard controls (see Control section above).

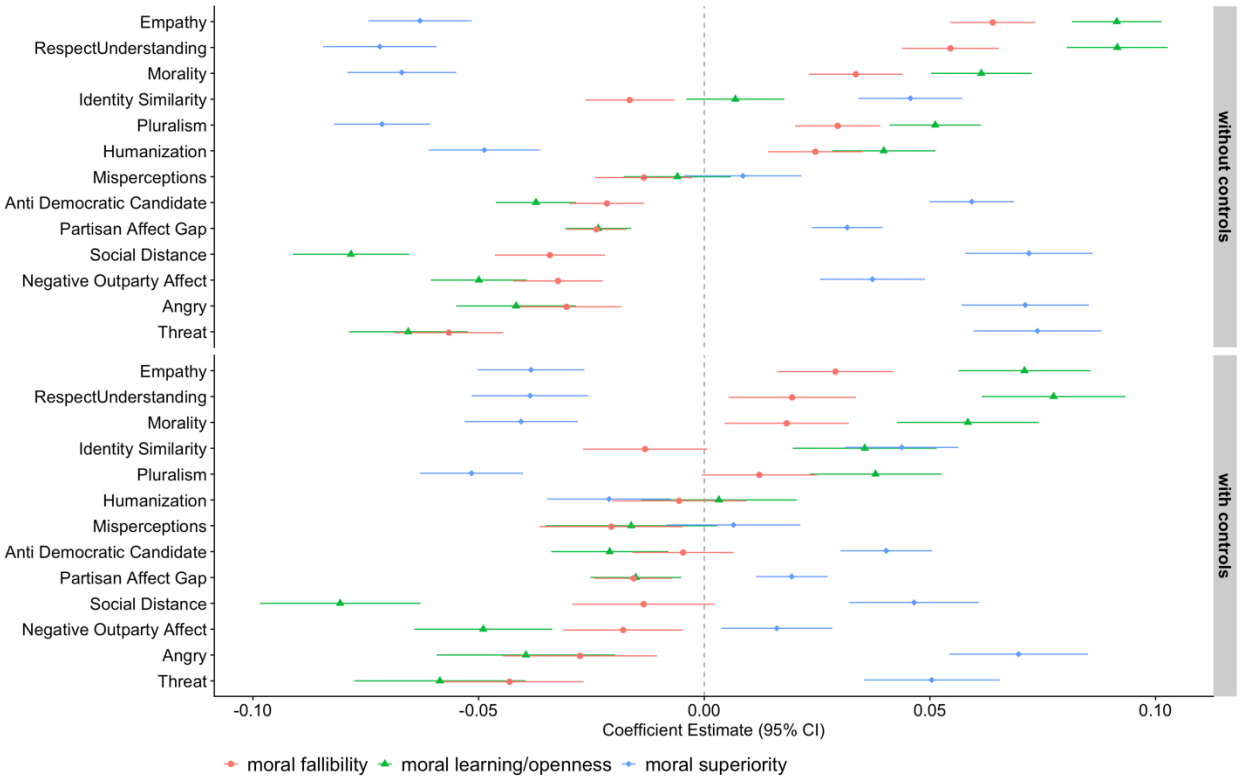
**Sample 4 Results: Conceptual Replication of Sample 3.** Three outcomes from Sample 3, misperception, partisan affect gap (or affective polarization), and anti-democratic attitudes were conceptually replicated in a national sample in Sample 4. Linear regression was used to estimate two models each, with and without controls, for the three outcomes. The results are in the lower part of Figure 2. Moral humility had negative relationships with all three outcomes (with and without controls).

**Sample 3 & 4 Subscale Result.** I also explored if the results observed in Sample 3 and 4 held or varied across the moral humility subscales. To this end, the models in Samples 3 and 4 were re-estimated using the moral humility subscales. The results are in Figures 3 and 4. Results observed for the whole scale largely held for all the subscales too. Some outcomes were less robust to inclusion of controls when using a particular subscale (such as pluralism when predicted by moral learning/openness, or misperceptions when predicted by moral superiority). However, the general direction of results across the outcomes for each subscale was consistent with that of the

whole scale. Moral superiority subscale is consistent with whole scale when it predicts outcomes in the opposite direction than the full scale i.e., more negative outcomes and less positive outcomes. The other two subscales should show results in same direction as the whole scale to be consistent. There was one exception. In Sample 5, higher moral fallibility predicted more support for anti-democratic norms.

**Figure 3**

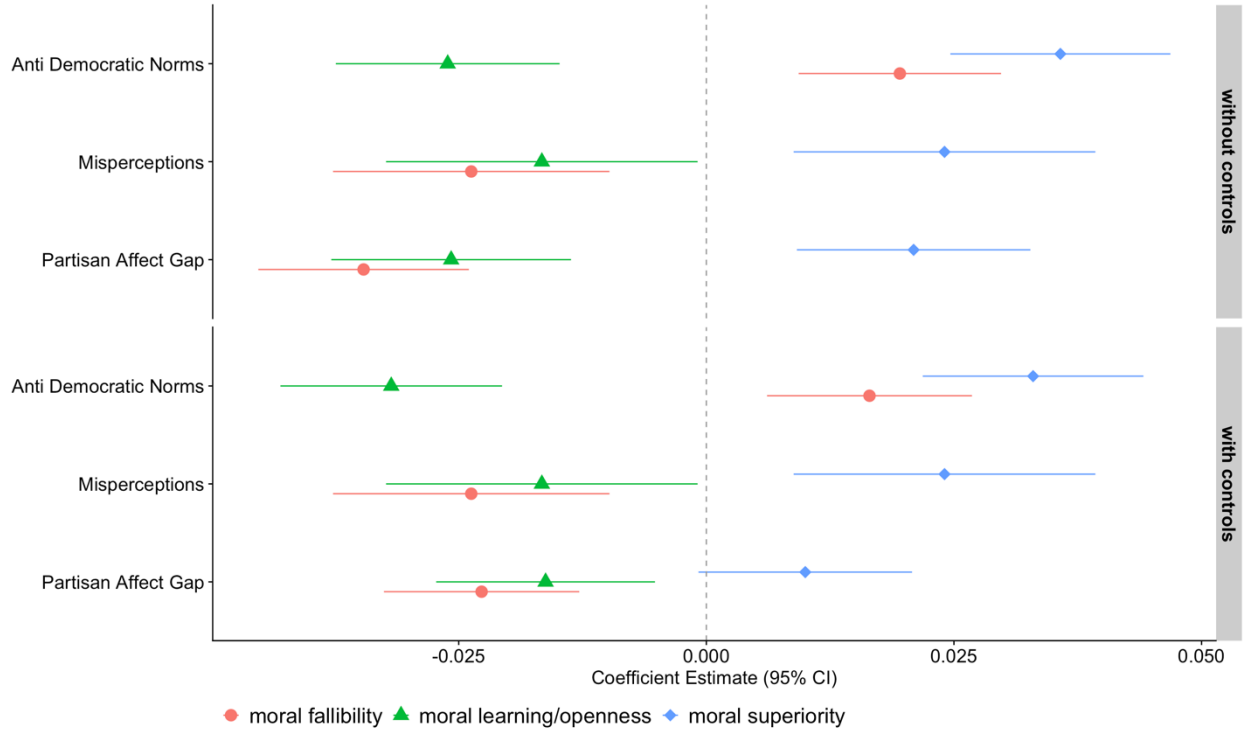
*Sample 3 Relationship between Moral Humility Subscales and Political Outcomes.*




*Note:* The green lines with triangles indicate where moral learning/openness was predictor; the orange lines with circles indicate the model where moral fallibility was the predictor; the blue lines with diamonds indicates the model where moral superiority was the predictor. Outcomes were recoded to range from 0-1. The controls in Sample 3 included the standard controls plus intellectual humility and moral relativism.

**Figure 4**

*Sample 4 Relationship between Moral Humility Subscales and Political Outcomes*




*Note:* The green lines with triangles indicate where moral learning/openness was predictor; the orange lines with circles indicate the model where moral fallibility was the predictor; the blue lines with diamonds indicate the model where moral superiority was the predictor. Outcomes were recoded to range from 0-1. The controls in Sample 4 only included the standard controls (see Control section above).

***Sample 4 and 5: Selective Exposure***

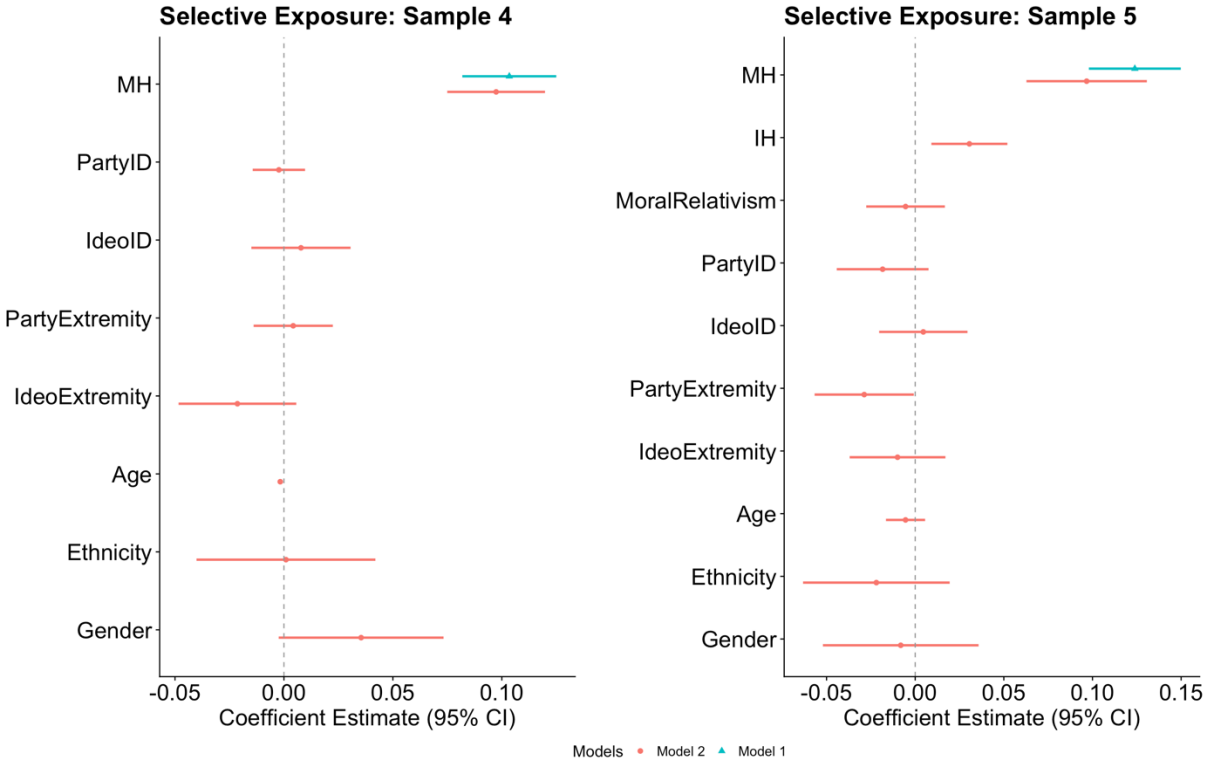
In both Samples 4 and 5, selective exposure was regressed on moral humility, both without and with controls. The analytic strategy was a little different in both samples given that Sample 5 was repeated measures design, unlike Sample 4. That is, in Sample 5, each participant completed selective exposure for three issues instead of just one.

In Sample 4, linear regression was used wherein selective exposure for one issue was regressed on the moral humility measure. Both the models, with and without controls, also included a dummy coded control for the issue that the participant had been randomly assigned to complete as part of the selective exposure task (described in methods). In Sample 5, as each participant completed selective exposure for three issues, multilevel regression analyses was used instead. Selective exposure was regressed on moral humility and issues were nested in persons. Random intercepts for both the issues and the persons were included.

Both samples found similar results for selective exposure (Figure 5). Higher moral humility was associated with more interest in cross-cutting exposure. This effect emerged over and above the control variables. Like before, models with each moral humility subscale were also estimated. The results are in Figure 6. Results observed for the whole scale largely held for all the subscales too.

**Figure 5**

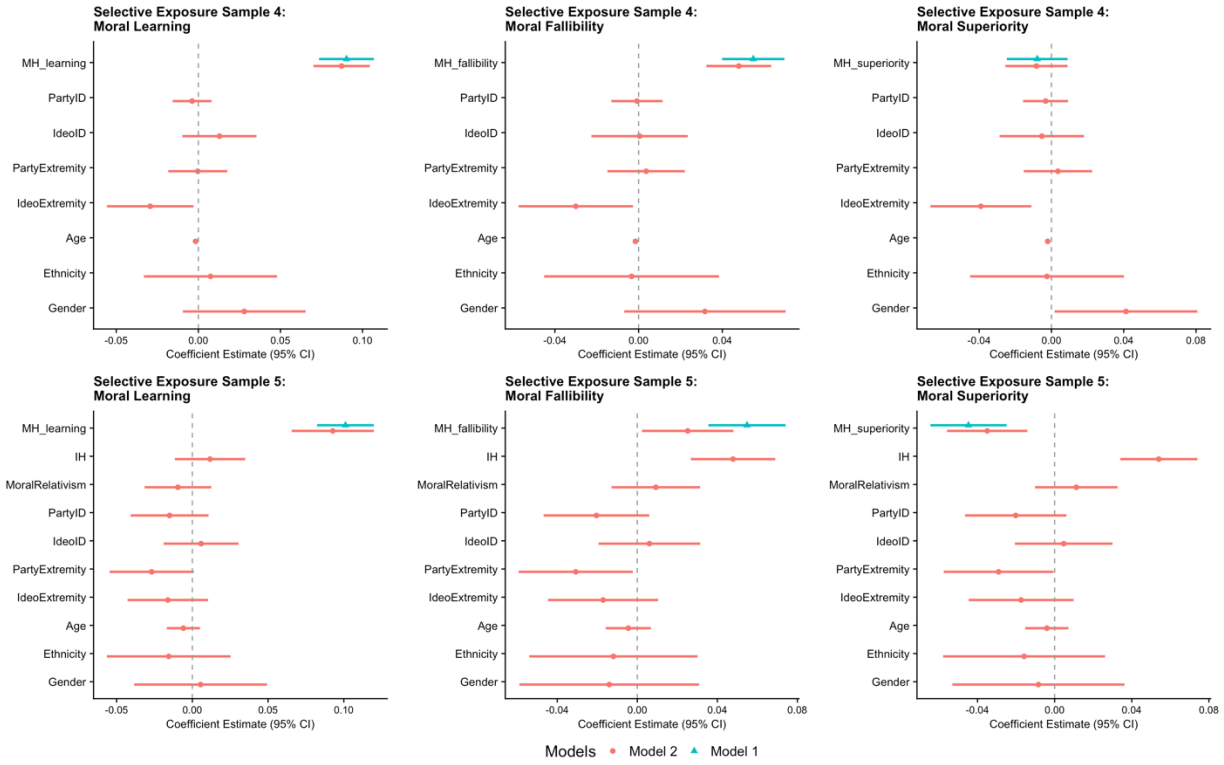
*Sample 4 and 5 Relationships between Moral Humility and Selective Exposure*




*Note:* The blue line with triangle indicates the model where moral humility was the only predictor; the orange lines with circles indicate the model where all covariates were also included. Outcomes were recoded to range from 0-1. Sample 4 was a national survey and hence has fewer controls. The coefficients for dummy codes for the issues in Sample 4 are hidden in the figure to keep the figure neat but were estimated in the model.

**Figure 6**

*Sample 4 & 5 Relationship between Moral Humility Subscales and Selective Exposure*




*Note:* The blue line with triangle indicates the model where moral humility was the only predictor; the orange lines with circles indicate the model where all covariates were also included. Outcomes were recoded to range from 0-1. Sample 4 was a national survey, and hence fewer controls were included. The coefficients for dummy codes for the issues in Sample 4 are hidden in the figure to keep the figure neat but were estimated in the model.

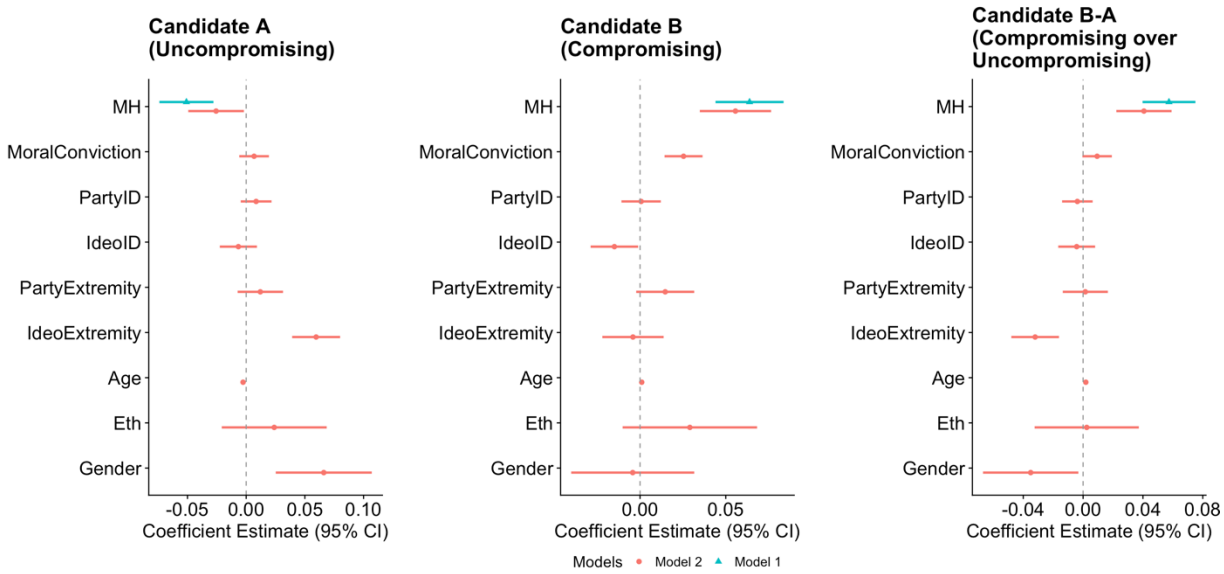
***Sample 6: Political Compromise***

The three political compromise outcomes (support for Candidate A, support for Candidate B, the relative support towards Candidate B over A) were each regressed on moral humility without and then with control variables.

Moral humility was associated with more support for Candidate B (the more compromising candidate), lesser support for Candidate A (the lesser compromising candidate), and more support of Candidate B over A, both with and without controls (Figure 7). The results for the subscales are in Figure 8. They show that the results observed for the whole scale largely held for all the subscales too. However, moral superiority more reliably predicted support for uncompromising candidate, whereas moral learning and fallibility more reliably predicted support for the compromising candidate.

**Figure 7**

*Sample 6 Relationship between Moral Humility and Political Compromise*




*Note:* The blue line with triangle indicates the model where moral humility was the only predictor; the orange lines with circles indicate the model where all covariates were also included. Outcomes were recoded to range from 0-1. The coefficients for dummy codes for the issues in Sample 6 are hidden in the figure to keep the figure neat but were estimated in the model.

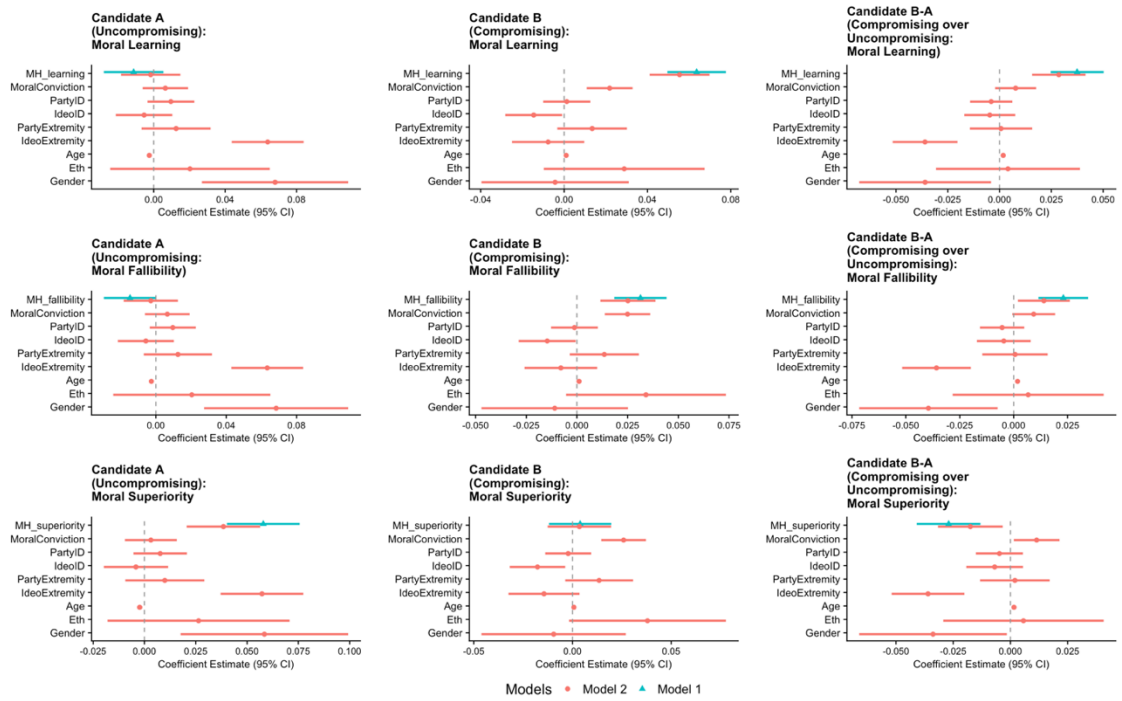
***Sample 7: Myside Sharing***

First, the political congruency of the ten news items (two for each of the five issues) that the participants saw as part of the task was determined for each participant. To this end, participants' position on the five issues were reverse coded where needed (i.e., on immigration question) such that lower values on all issue indicated more conservative position and higher values indicated more liberal position. If the participant had a response <50 on an issue, then the conservative news item on the issue was considered as politically congruent and the liberal news item as politically incongruent. If the participant had a response ≥50 on an issue, then liberal news items were considered congruent and conservative news items incongruent. Two contrast codes indexing news item congruence were created for analyses wherein incongruent news stories were treated as reference. The two contrast codes were congruent (congruent = 0.5, other = -0.5) and neutral (neutral = 0.5, other = -0.5).

Multilevel regression models were used for the final analysis with random intercepts for participants and news item. Moral humility was grand-mean centered. In the first set of models, willingness to share the news item was first regressed on congruent dummy code, neutral dummy code, and congruent dummy's interactions with moral humility. In the second set of models, control variables were included.

**Figure 8**

*Sample 6 Relationship between Moral Humility Subscales and Political Compromise*




*Note:* The blue line with triangle indicates the model where moral humility was the only predictor; the orange lines with circles indicate the model where all covariates were also included. Outcomes were recoded to range from 0-1. The coefficients for dummy codes for the issues in Sample 6 are hidden in the figure to keep the figure neat but were estimated in the model.

Results showed that people were indeed more likely to share congruent news items more than incongruent news items ($b$ = 0.71, $SE$ = 0.03, $t$ = 23.54, $p$ <.001). The same was observed for neutral new items ($b$ = 0.72, $SE$ = 0.17, $t$ = 4.06, $p$ <.001). However, there wasn't a significant interaction of moral humility with congruent dummy, contrary to what was predicted ($b$ = -0.03, $SE$ = 0.02, $t$ = -1.33, $p$ = 0.18). Moral humility did not have a main effect either; it wasn't associated with levels of news sharing ($b$ = 0.03, $SE$ = 0.04, $t$ = 0.8, $p$ = 0.40). Results for the subscale were largely consistent with the results of the whole scale. There wasn't a significant interaction of moral learning/openness with congruent dummy ($b$ = 0.01, $SE$ = 0.02, $t$ = 0.8, $p$ = 0.41). Moral fallibility ($b$ = -0.04, $SE$ = 0.02, $t$ = -1.80, $p$ = 0.06) and moral superiority ($b$ = 0.04, $SE$ = 0.02, $t$ = 1.80, $p$ = 0.07) had marginal $p$ values and small effects in the direction of predicted effects (below the threshold of what our study was powered to detect at 90%). The results for the scale and subscale held with inclusion of control variables. Overall, the results suggested

that moral humility and its various facets didn't significantly temper the myside sharing effect. If there is indeed an effect in the predicted direction in case of moral fallibility and superiority, the effect might be small.

***Incremental Validity Robustness Check***

Moral humility significantly predicted polarization and related outcomes over and above important controls like moral relativism and intellectual humility. This provided support for its incremental validity. However, recently some research has advised against using multiple regression solely to establish evidence of incremental due to the likelihood of Type I error from measurement unreliability (Westfall & Yarkoni, 2016). The authors instead recommend methods like structural equation modeling (SEM) which accounts for measurement unreliability. Given that establishing moral humility's incremental validity over closely related constructs like moral relativism and intellectual humility was one of the central aims in this project, I checked for the robustness of the incremental validity results in Sample 3 and 5 (these had the two controls in analyses) using this recommended SEM approach. Instead of computing the mean score for moral humility, moral relativism, and intellectual humility and then entering these mean scores as simultaneous predictors in a multiple regression, I specified the measurement models for these three constructs. Thus, moral humility (hierarchical model with three factors), moral relativism (one factor model), and intellectual humility (one factor model) were entered as latent factors as part of the structural regression model. Accordingly, the measurement error became an explicit part of the full model. I tested the models with the control variables in Sample 3 using this approach (see SOM). Essentially, I replicated moral humility's incremental validity over the control variables, and notably over and above moral relativism and intellectual humility. Interestingly, most significant effects of intellectual humility that were observed before no longer held when using structural equation modeling. Further, the effect sizes for moral humility also became larger.

***Predictive Validity Effect Size***

I computed standardized regression coefficients for all models across studies (see SOM). A summary of the moral humility effect size from all models with and without control are in Table 11. For comparison, I also present effect size from the models with controls for intellectual humility — one of the constructs closest to moral humility which has also been investigated in the polarization context before, and political (partisan and ideological) extremity — the construct that has been theorized to be central to polarization. Compared to both intellectual humility and political extremity, moral humility showed a stronger effect. Notably, these moral humility effects were even

stronger (median $\beta = 0.46$) when SEM was used for analyses which accounts for measurement unreliability (see incremental validity robustness section above).

The summary statistics (Table 11) show that the average effect of moral humility ($\beta \sim 0.20$) was small according to Cohen's standards. According to different standards, an effect of 0.2 is medium (Funder & Ozer, 2019) while an effect of 0.1 is small. According to these guidelines, the effects for moral humility were on an average, medium-sized, and those of intellectual and political extremity were small-sized. In sum, the effect size of moral humility being comparable (or larger) to political extremity suggests that moral humility is perhaps of substantive importance in understanding polarization.

**Table 11**

*Predictive Validity Effect Size Summary for Moral Humility, Intellectual Humility, and Political Extremity*

| | **MH (Without Control)** | **MH (With Control)** | **IH (With Control)** | **Party Extremity (With Control)** | **Ideological Extremity (With Control)** |
|---|---|---|---|---|---|
| Mean | 0.23 | 0.20 | 0.09 | 0.12 | 0.13 |
| SD | 0.09 | 0.08 | 0.06 | 0.09 | 0.07 |
| Median | 0.23 | 0.19 | 0.11 | 0.09 | 0.14 |
| Min | 0.07 | 0.07 | 0.02 | 0.01 | 0.02 |
| Max | 0.42 | 0.34 | 0.19 | 0.28 | 0.23 |

**Summary of Polarization Results**

The broader aim of the analyses conducted in this section was to test if in a moralized and conflictual context, moral humility could attenuate its negative aspects, such as having low opinion of the outgroup, antagonism and derogation towards outgroup, rigidity in one's own views, rejection of compromise and contact, and adoption of morally questionable means.

In line with these expectations, in the context of political polarization in the US, I found moral humility to negatively predict a range of polarization and other associated outcomes across studies. Moral humility was associated with more positive opinion towards political outgroup such as higher perceptions of outgroup morality, lower perceptions of outgroup threat, and lower misperception of the outgroup's hostility towards their ingroup. Moral humility was also associated with lesser antagonistic feelings towards political outgroups such lower negative affect and lower anger, as well as lower gap between negative/positive feelings towards political ingroups and outgroups. It was further associated with a range of other-oriented outcomes, such as more empathy, more respect, and more understanding towards political outgroup and their perspectives. Similarly, it was associated with less

rigidity and more openness, such as expressing more willingness to learn from disagreeing others and their views, more willingness to engage in political compromise, and not very strong attachments to the political ingroup identity. Finally, it was associated with lower support for morally questionable means such as support for anti-democratic and anti-pluralistic actions. Thus, overall people higher in moral humility demonized the political outgroup less, had more positive, respectful and open orientation towards the political outgroup and towards opposing political viewpoints, and were more committed to norms and practices accommodating diverse values, perspectives and interests.

These studies taken together provided evidence for moral humility's criteria and predictive validity in a moralized context. Further, most of these effects were observed over and above the effects of closely related psychological constructs such as intellectual humility, moral relativism, and political extremity. This thus provided evidence for moral humility's incremental validity and suggested that it has distinct value in explaining moralized conflicts. Further, the effect of moral humility on polarization outcomes, although modest, was comparable (and even stronger) to the effects of political extremity, an explanatory variable considered very central in polarization literature. This suggests that moral humility is perhaps of substantive importance.

However, there were a few findings that did not fully align with expectations. First, moral humility did not predict lower myside sharing on social media, and the effects of moral humility facets which showed effects in prediction-consistent direction were small and marginal. There could be a few reasons for that. Political conversations are a small proportion of total conversation on social media and a very small number of social media users (~9%) produce the most political content (Pew Research Center, 2019, 2021). This suggests that most people do not engage in political sharing on social media. This is supported by our own data which found the median sharing to be 2 (indicating "Unlikely" on a scale of 6, with mean of $M = 2.5$ with $SD = 1.5$). Thus, maybe moral humility might not have much variation to explain, or other factors might be more important in explaining political sharing on social media.

Second, in one of the samples (Sample 4), moral fallibility subscale was positively associated with more support for anti-democratic actions. This was not expected a priori, but aligns with the results found in previous section, vis-à-vis moral fallibility's small positive associations with psychopathy, which had suggested that moral fallibility might be capturing some people who accept engaging in or supporting morally questionably acts *and* also accept their immoral inclinations as moral limitations. In a similar vein, the results for anti-democratic attitudes

suggest that people who support anti-democratic means might also be willing to accept that they might be morally flawed sometimes. However, there is a caveat here — this positive relationship was only found in one sample and not the other (Sample 3) suggesting it would need to be replicated and investigated further before strong conclusions can be drawn.

Otherwise, the results for subscales largely held in direction consistent with the overall scale. Some effects were weaker for a particular subscale or were sometimes less robust to controls, suggesting that different facets of moral humility might be sometimes playing a stronger role in regard to some outcomes. For example, support for the less compromising candidate (Sample 6), support for anti-democratic candidate (Sample 3) and lower support for pluralistic norms (Sample 3) was more reliably and robustly predicted by moral superiority. Or higher willingness to cross-cutting exposure (Sample 4, 5) was more reliably and robustly predicted by moral fallibility and moral openness/learning. Thus, the different facets of moral humility play unique role in explaining outcomes while being largely consistent with results of the whole scale, supporting the view of moral humility construct as one construct with different and unique yet correlated factors. Taken together the results indicate that moral humility predicts outcomes in sensible and expected ways across the scale and the subscales, and has unique explanatory power over other constructs of substantive importance.

## Chapter III: Intervention Development and Causal Assessment

The studies conducted so far used cross-sectional designs. Further, moral humility was measured like a trait, i.e., as a stable dispositional aspect of a person. In the next step of my research program on moral humility, and as part of the new studies in the dissertation involving the committee, I conducted two experiments to take the study of moral humility in a causal direction, as well as investigate if moral humility can be studied as a state. That is, I tested if and how moral humility can be changed. The studies attempted to change moral humility using newly designed interventions that manipulated aspects of moral humility like fallibility, openness, learning, and other-orientedness in order to move people's moral humility. Further, the studies also tested if these changes have downstream effects and are associated with reduced levels of polarization outcomes. Together, these studies thus attempted to find ways that moral humility can be changed and establish the causal effect of moral humility on political outcomes. These studies were the first experimental studies that I know of that attempted to do this.

First, I developed five moral humility interventions, details of which are described first below. I then pilot tested these and subsequently ran two experiments, testing these interventions individually (Experiment 1) and together (Experiment 2). The Pilot, Experiment 1, and Experiment 2 are described below in separate sections after describing the five interventions.

### Transparency and Openness

Both the main experiments' study materials, design, hypotheses, and planned analyses were preregistered and can be found at OSF. All studies were approved by the Michigan State University Institutional Review Board.

### Moral Humility Interventions

Five interventions or treatments were developed to increase moral humility (see SOM for the interventions). These interventions each had a vignette that the participants would read, which talked about some moral ideas that I believed might move moral humility. The vignettes were presented along with congruent pictures and interactive questions based on the vignette. These interactive questions were not intended to be the outcome variables in main experiments, but were a feature of the treatments themselves, included to make the task of reading these vignettes more engaging. There was also a control vignette with pictures and interactive questions on a topic (artificial intelligence) unrelated to morality, to be used in the control condition in the main experiments.

The five moral humility vignettes aimed to induce moral humility by manipulating or making salient its various attributes — moral fallibility, moral openness/learning, and moral superiority. Each vignette invoked one or

more of these attributes. It might be argued that a better design would have different vignettes targeting different moral humility attributes separately than employing a mix of them in the vignettes. This systematic approach might be useful in future work. At this nascent stage where no moral humility manipulations exist in the literature, the aim was to adopt more of a proof-of-concept approach to see if moral humility can be moved in principle and what kind of ideas broadly work well as moral humility manipulations. To that end, a more general approach was adopted to designing manipulations, using any ideas that might capture and move moral humility more generally. These ideas were thus a mix of moral humility attributes that I believe might move moral humility.

In a similar vein, a secondary justification for adopting a general approach of using a mix of moral humility attributes in the intervention vignettes is that previous work testing multiple interventions to reduce polarization (Voelkel et al., 2023) found that multifactorial interventions (i.e., treatments that employed multiple strategies) were more effective. Thus, at a proof-of-concept stage where the aim was to move moral humility, employing strategies that maximize effectiveness rather than isolate the precise causal factors might be preferable. Notably, such an approach has a long history in psychological intervention literature (Wilson et al., 2010; Rozin, 2006) wherein identifying strategies that can shift an important outcome is given preponderance over identifying or isolating exact theoretical mechanism or causal factors that can move the outcome of interest (example of such recent works include Broockman & Kalla, 2016; Voelkel et al., 2023; for a review see Paluck et al., 2021) . The former is captured in research that is called “problem-oriented” whereas the latter in research that is called “process-oriented” wherein studying the phenomenon versus studying the process is respectively emphasized — the former is recognized as a suitable approach especially during initial investigations of a problem or phenomenon where researchers are still exploring “is” questions (e.g., what is the phenomenon?) (Wilson et al., 2010). A recent example of problem-oriented research is Broockman and Kalla (2016) where researchers aimed to reduce transgender prejudice. The *problem* here was prejudice reduction. To that end, the study had canvassers going door-to-door talking to people about transgender laws, including showing them a video of opposing views on transgender issues, asking them to talk about a time when they were judged for being different, asking them to engage in perspective taking with transgender experiences, and asking for report on whether the whole exercise changed their minds. Thus, the study employed a variety of strategies to the end of prejudice reduction instead of focusing on isolating and identifying individual causal factors. The idea behind the experimental studies testing moral humility interventions was also somewhat similar.

It is important to note that the interventions were not designed with the aim of producing large or lasting effects (such as longitudinal treatments), although I do present combinations of the five moral humility treatments in Experiment 2 to test if doing so would produce bigger effects than when the treatments are presented individually. In any case, they were light touch interventions delivered in a single online exposure with the intention of temporarily boosting a state of humility toward one's own and others' morality.

The arguments or ideas presented as part of the vignettes were based on moral psychological and philosophical work (Baumeister, 1999; Zimbardo, 2004; MacAskill et al., 2020; Williams, 2015; Tersman, 2022; Cole, 2023). All five moral humility vignettes as well as the one control vignette were largely matched for length and style of presentation. The participants were asked to read these vignettes as part of a study that aims to understand people's thoughts about various philosophical ideas. See SOM for the exact wording and presentation of the vignettes.

The *first vignette* highlighted our psychological tendency towards bias and overconfidence. It pointed out how our views about right and wrong are often biased due to things like our personality, background, and social relationships, how we are usually blind to our biases and fail to correct them, and on top of it are overconfident in our views. It suggests that a more accurate understanding of what is morally apt may be easier if people lowered their confidence in their own moral standpoints and had more openness to learning from people with different moral standpoints. This vignette thus highlighted the facets of moral fallibility, moral learning and openness, and moral superiority.

The *second vignette* highlighted the difficulty of being ethical, such as figuring out the right moral view or action, living up to our values, developing a good moral character, weighing competing moral ideas and values accurately, understanding the complexities of different moral situations, and accounting the needs and perspective of all moral stakeholders. It is suggested to the reader that such complexities mean that our moral judgments might be prone to error, and that making the right moral choice or being morally upstanding can be difficult. It proposed that being moral is thus a learning process. This vignette thus highlighted the facets of moral fallibility, and moral learning and openness.

The *third vignette* highlighted ordinary people's capacity of evil. It made salient the fact that it is not especially evil people who commit moral wrongdoings, such as those in Nazi Germany, Maoist China, and the Soviet Union, but rather ordinary people like us who deceive themselves of doing a noble deed. It suggests that we

should therefore be mindful of our ability to commit horrible misdeeds. This vignette thus highlights the facets of moral fallibility, moral openness, and moral superiority.

The *fourth vignette* highlighted humankinds' morally imperfect past and suggested that like every generation before us, we are also likely blind to our moral limitations and unknowingly participate in terrible moral transgressions. It suggested that one way to do a better job of identifying our moral mistakes and correcting them would be to set up a social environment supporting exchange of ideas so that different people could combine their knowledge and abilities to identify and correct our moral blind spots. This vignette thus highlighted the facets of moral fallibility, moral learning and openness, and moral superiority.

The *fifth vignette* highlighted the existence of moral disagreements amongst equally rational, morally sensitive, well-informed, and well-intentioned people. It suggests that such disagreement might indicate that other people might have useful moral perspectives that might help us reach accurate understanding in moral matters. It also highlighted that people who we morally disagree with often have other moral strengths. This vignette thus highlighted the facets of moral fallibility, moral learning and openness, and moral superiority.

The *control vignette* explored whether AI (artificial intelligence) can think and understand as humans. It presents John Searle's Chinese Room thought experiment to the reader. This vignette was designed with the intention of not evoking any content overlapping with moral humility and its facets, while still being engaging.

**Possible Theoretical Mechanism**. Although identifying and isolating the precise theoretical or psychological mechanisms undergirding the interventions was not a priority, I speculated that some of the following factors and mechanism will perhaps drive the interventions' effects (if any). First, the interventions might increase the *salience* of certain aspects undergirding moral humility. For instance, reading about the difficulty of being ethical, or about our tendencies towards bias or overconfidence, or the atrocities committed by our ancestors might *activate* the attribute of moral fallibility thereby making our moral limitations salient.

The interventions might also be *educational*. The interventions inform and make people aware of our moral flaws and limitations such as our biases and overconfidence, or of the ordinariness of people who committed well known historical atrocities. Being educated about these might be a morally humbling experience. These might therefore move people to recognize the necessity of moral humility.

Along that line, the interventions might implicitly suggest various aspects of moral humility as solutions to the problems raised in the interventions, such as increasing moral learning/openness or reducing moral superiority as

a solution to our biases and overconfidence. Similarly, suggesting recognition of one's moral fallibility as a solution to our ability to commit immoral deeds. The interventions therefore implicitly aim to *increase the value* of moral humility to the reader and presents it as an important virtue by framing it as a solution to some problems (moral or otherwise) that people might relate to. The idea is similar to when a political party first makes a problem salient to the voters and then presents voting for their candidate or party as a solution to fixing that problem. Or an advertisement from a company makes the consumer salient of an issue or need and then presents their product as a solution to addressing that issue or need. Such techniques have been called by different names in the persuasion literature, like problem–reaction–solution (PRS) framing (Drinkwater et al., 2018), or demand or need creation strategy (Priem et al., 2018). The interventions thus are in a way intended to create a "demand" for moral humility.

The interventions might also evoke certain *emotions* that might aid in boosting moral humility — for instance highlighting the moral wrongdoing of others like them or their ancestors might evoke shame, embarrassment, or fear which might induce humility about morality. However, it is important to note that past work on humility interventions suggests that exposure to negative self-information is not conducive to inducing humility by itself unless accompanied by self-affirmation, given that self-affirmation helps put people in a less defensive state of mind and facilitates acknowledgement of one's limitations (Ruberton et al., 2016). Along these lines, in order to not make the reader feel attacked, despondent, or nihilistic when reading about moral flaws and limitations, the interventions incorporated solutions when presenting problems.

All the mechanisms noted here are often used in psychological interventions (e.g., for salience manipulation see Burke et al., 2010, Van Tongeren et al., 2016; for educational manipulation see Paluck et al., 2021; for problem-solution manipulation Tannenbaum et al., 2015; for emotion manipulations see Ruberton et al., 2016). I conjectured that some or all of these mechanisms might be at play in the moral humility interventions.

## Pilot Study

Both experiments were preceded by a small Pilot study wherein I tested the five moral humility interventions that I developed and are described above. The aim was to assess if the five interventions to be used in the main experiments were well received by the participants, with the intention of making changes in the interventions accordingly before the main experiments.

### Dataset and Participants

The Pilot study was conducted on Prolific. Prolific is an online service that facilitates the crowdsourcing of research participants (Douglas et al., 2023; Peer et al., 2022). The participants on Prolific are more diverse than the average college sample than other crowdsourcing platforms like MTurk (Palan & Schitter, 2018) and provide high quality data (Peer et al., 2022). The Pilot was opened to 50 participants aged over 18 and evenly recruited from those who self-identified as Democrats and Republicans in Prolific's prescreening to make sure the interventions are well-received by participants of diverse ideological leaning. Participants were paid $4 for doing the study (~20 minutes). The Pilot did not have treatment or control conditions. All participants read all five moral humility treatments and completed measures that assessed their impressions of the treatments. Demographics was not collected in the Pilot.

### Materials and Procedure

#### *Moral Humility Interventions*

All participants read and interacted with all five moral humility interventions described in the previous section and then answered some questions (described next). The control vignette was not included in the pilot-testing to avoid making the Pilot too long and was directly used in the main experiments.

#### *Outcome Variables*

After reading and interacting with the five treatments in the Pilot, the participants answered outcome measures assessing their evaluation of what they just read. The outcome measures are summarized in Table 12 (table also contains a summary of measures included in the main experiments).

**Table 12**

*Post-treatment outcomes and pre-treatment covariates included in different studies*

| **Outcomes (Post-Treatment)** | **Pilot** | **Experiment 1** | **Experiment 2** |
|---|---|---|---|
| Moral Humility | | ✓ | ✓ |
| Intellectual Humility | | ✓ | |
| Moral Relativism | | ✓ | |
| Political Sectarianism | | ✓ | ✓ |
| Selective Exposure | | ✓ | ✓ |
| Political Compromise | | ✓ | |
| Epistemic Emotions | ✓ | ✓ | |
| Message Quality | ✓ | | |
| Thoughts | ✓ | | |
| **Covariates (Pre-Treatment)** | **Pilot** | **Experiment 1** | **Experiment 2** |
| Moral Humility | | ✓ | ✓ |
| Political Sectarianism | | ✓ | ✓ |
| Selective Exposure | | | ✓ |

**Message Quality**: The perceived quality of the vignettes was assessed along two indicators: (i) the ease of comprehensibility, i.e., if participants found the vignettes easy to read, and (ii) interestingness, i.e., if participants found the vignettes engaging to read. Participants answered these after reading each vignette. These were intended to help provide valuable information to evaluate the weaknesses of the vignettes so that changes could be made if needed. On a 7-point scale from *1 (extremely disagree)* to *7 (extremely agree)*, participants answered the following 2 items after each vignette — “The text was easy to read.” (to assess ease of comprehension), and "I found the content of the text interesting” (to assess the interestingness).

**Emotions.** Participants also reported the strength of various epistemic emotions they experienced during the reading of the vignettes on Epistemically-Related Emotion Scales (Pekrun et al., 2017). The scale is meant to capture affective states that occur during cognitive activities involving acquisition or generation of knowledge, such as surprise, curiosity, enjoyment, confusion, anxiety, frustration, and boredom. Assessing these epistemic emotions in the Pilot provided information on whether the treatments were well-received by participants, which were again intended to be used to make changes to the treatments before the main experiments if needed. The fourteen items

capturing the seven core emotions (two items each) were surprised, amazed, curious, interested, excited, happy, confused, muddled, anxious, nervous, frustrated, irritated, bored, and monotonous, and were answered on a 7-point scale *(1=Not at All, 4=Moderate, 7=Very Strong).* Unlike message quality, these were not assessed after each vignette, but at the end of reading all vignettes to assess participants' overall experience with the treatments.

**Thoughts.** After each vignette, participants were also provided with an open-ended box where they were asked to briefly respond to "What did you think of the idea you just read? What came to your mind when you read it? Did you like or dislike anything?" These open-ended responses were used to get a more detailed picture of what the experience of the participants was when reading the vignettes, and if they negatively responded to some ideas.

**Analysis and Results**

The aim of the Pilot study was to examine if the five treatments or interventions that would be used in Experiment 1 and 2 were received well by the participants.

***Ease of comprehension***

Descriptive statistics for the ease of comprehension for each of five treatments was estimated. Results are in Table 13 and presented in Figure 9. Results indicated that all treatments were easy to understand, with mean perceived easiness being above midpoint for all treatments. Further, most ratings clustered above the midpoint, indicating that most participants generally found the vignettes easy to read.

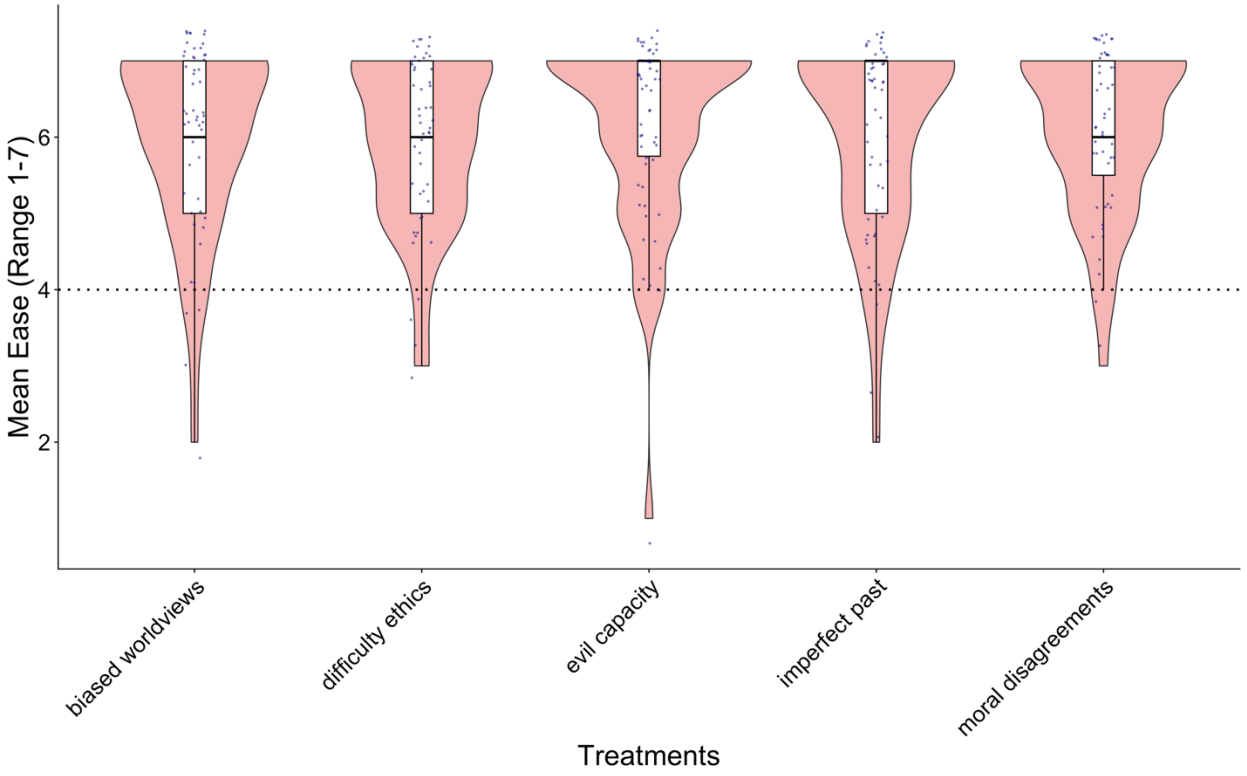
**Table 13**

*Summary statistics for ease of comprehension across the five treatments*

| | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|
| biased worldviews | 6 | 6 | 1.19 | 2 | 7 |
| difficulty ethics | 5.96 | 6 | 1.09 | 3 | 7 |
| evil capacity | 6.12 | 7 | 1.22 | 1 | 7 |
| imperfect past | 6.02 | 7 | 1.24 | 2 | 7 |
| moral disagreements | 6.1 | 6 | 1.02 | 3 | 7 |

**Figure 9**

*Distribution of the perceived ease ratings across the five treatments*




***Interestingness***

Descriptive statistics for the perceived interestingness of each of five treatments was estimated. Results are in Table 14 and presented in Figure 10. Results indicated that all five treatments were perceived to be largely interesting, with mean perceived interestingness above midpoint for all treatments. Further, most ratings clustered above the midpoint, indicating most participants generally found the vignettes interesting to read.

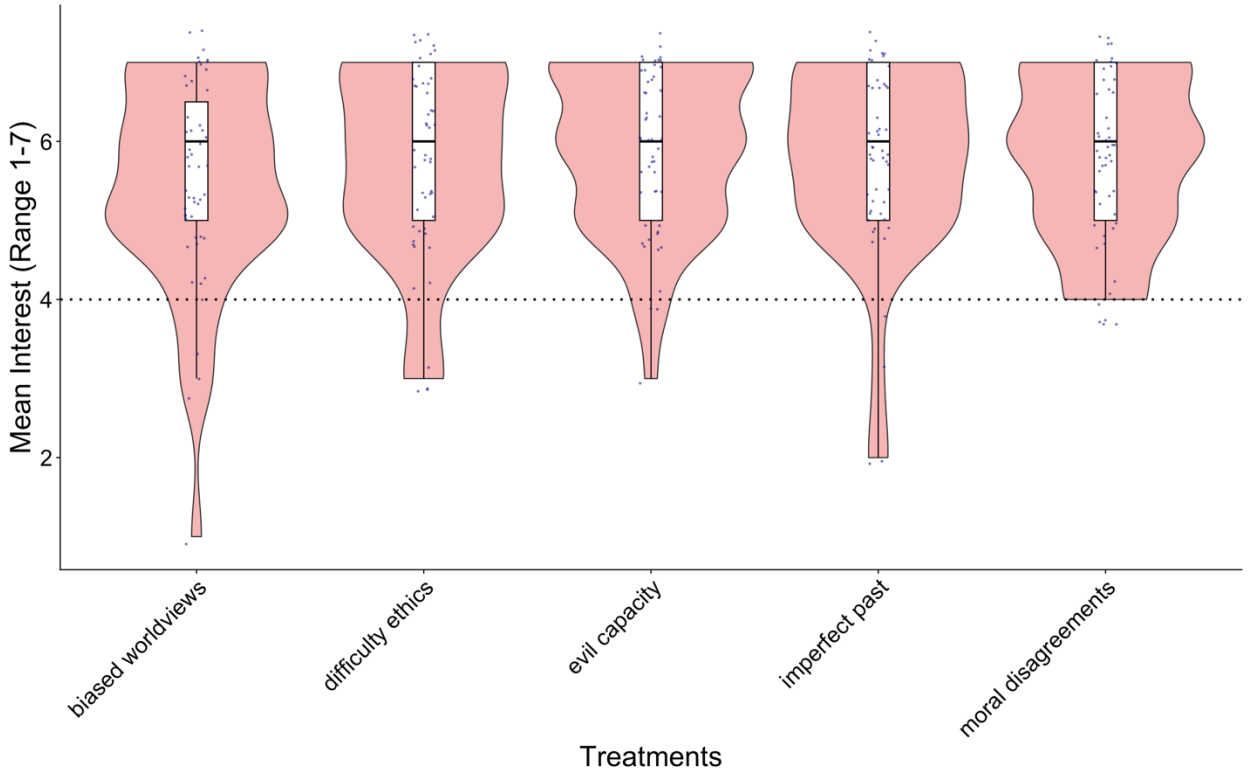
**Table 14**

*Summary statistics for interestingness across the five treatments*

| | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|
| biased worldviews | 5.49 | 6 | 1.3 | 1 | 7 |
| difficulty ethics | 5.7 | 6 | 1.2 | 3 | 7 |
| evil capacity | 5.92 | 6 | 1.02 | 3 | 7 |
| imperfect past | 5.76 | 6 | 1.21 | 2 | 7 |
| moral disagreements | 5.78 | 6 | 1.03 | 4 | 7 |

**Figure 10**

*Distribution of the perceived interestingness ratings across the five treatments*




***Emotions***

Descriptive statistics for the fourteen epistemic emotions (two items each for seven core emotions) was estimated. Unlike other indicators, these were not reported for each treatment but asked at the end of all five treatments to get an overall sense of the emotions evoked by the treatments. Results are in Table 15 and presented in Figure 11. The results indicate that interest and curiosity were the most strongly experienced emotions, with mean scores above midpoint for both. This suggests that overall reading the treatments was interesting and curiosity-inducing for participants. Mean level of emotions that would indicate a negative experience with the treatments such as muddled, monotonous, irritated, confused, frustrated, and bored were all below midpoint.

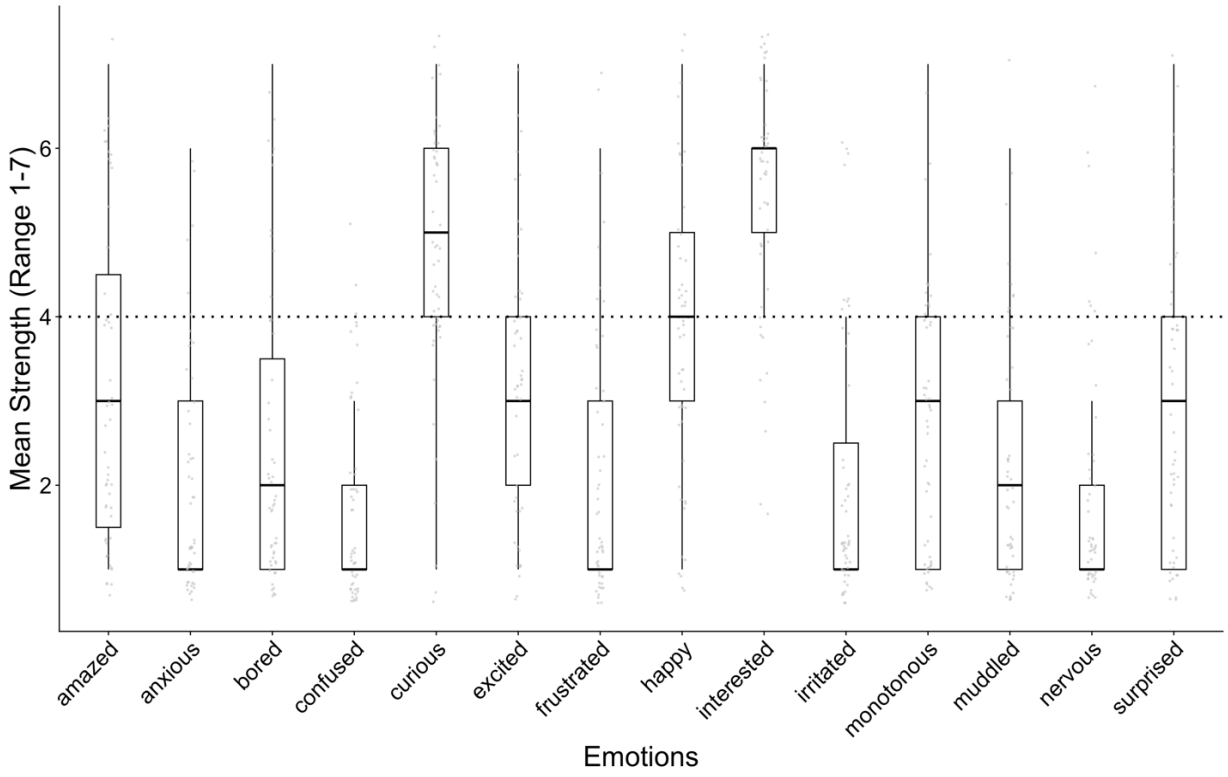
**Table 15**

*Summary statistics for epistemic emotions*

| | Mean | Median | SD | Min | Max |
|---|---|---|---|---|---|
| amazed | 3.12 | 3 | 1.91 | 1 | 7 |
| anxious | 2.08 | 1 | 1.45 | 1 | 6 |
| bored | 2.49 | 2 | 1.79 | 1 | 7 |
| confused | 1.84 | 1 | 1.14 | 1 | 5 |
| curious | 4.75 | 5 | 1.56 | 1 | 7 |
| excited | 3.22 | 3 | 1.64 | 1 | 7 |
| frustrated | 2.25 | 1 | 1.67 | 1 | 7 |
| happy | 3.75 | 4 | 1.72 | 1 | 7 |
| interested | 5.39 | 6 | 1.37 | 2 | 7 |
| irritated | 2.08 | 1 | 1.59 | 1 | 6 |
| monotonous | 2.69 | 3 | 1.54 | 1 | 7 |
| muddled | 2.2 | 2 | 1.52 | 1 | 7 |
| nervous | 1.96 | 1 | 1.55 | 1 | 7 |
| surprised | 3.04 | 3 | 1.83 | 1 | 7 |

**Figure 11**

*Distribution of the strength of epistemic emotions*



### *Thoughts*

The open-ended thoughts that participants reported at the end of reading each treatment were also analyzed, wherein their thoughts were categorized as either positive, negative, mixed, or neutral. Results of the subjective coding suggested that no treatment was uniquely negatively perceived by the participants to a concerning extent. There were ~5-7 negative responses for each treatment out of a total of 51-52 responses. Results are in Table 16.

**Table 16**

*Summary of qualitative coding of open-ended thoughts provided at end of each treatment*

| | Positive/ Agreement | Negative/ Disagreement | Mixed | Neutral/ Non-answers |
|---|---|---|---|---|
| difficulty ethics | 33 | 6 | 1 | 11 |
| biased worldviews | 38 | 5 | 5 | 4 |
| imperfect past | 27 | 6 | 5 | 13 |
| evil capacity | 30 | 7 | 3 | 12 |
| moral disagreements | 35 | 6 | 3 | 7 |

*Note:* Positive/Agreement or Negative/Disagreement is conceptualized broadly. Former includes responses where participants are echoing thoughts in the text and latter includes responses where participants are expressing a counter thought to the text presented.

### Pilot Summary

The results of the pilot indicated that the text in all treatments was easy to understand, perceived as engaging, and did not evoke negative emotions or thoughts. Taken together, the indicators suggested that all five treatments were perceived positively and received well by the participants, and thus these experimental materials were well-suited for use in the main experiments without modification.

## Experiment 1

Experiment 1 tested the five moral humility interventions or treatments individually. The experimental setup thus had five experimental conditions tested against a control condition. The experiment used a pre-post design to increase precision (Clifford et al., 2021) wherein outcomes were measured before the treatment as well as after the treatment. The idea behind these designs is that when the pre-treatment outcome measure is correlated with the post-treatment outcome measure, it is possible to increase the precision of the estimates and the power of the study. This study aimed to (i) provide a proof-of-concept for moral humility being amenable to change, ii) to examine which treatments work best to increase moral humility, and (iii) assess the causal relationship between moral humility and polarization. All preregistered study details can be found at OSF.

**Dataset and Participants**

Experiment 1 was conducted on Prolific. The study was opened to 2700 US participants aged over 18. The sample details are in Table 17. Participants were paid $2.60 for doing the study (~13 minutes) which is $12/hr. Experiment 1 had five experimental conditions and one control conditions. A sample of 2700 people (~ 450 in each condition) provided good statistical power (~80-85%) in a pre-post experimental design to detect a small effect size between each of the five treatment conditions with control condition ($d \sim 0.15$) where the pre and post treatment outcomes are correlated at .65 (for reference, minimum split-half reliability of moral humility in previous work was ~0.8). If and when the pre- and post-treatment outcomes' correlations are higher, the achieved statistical power would be higher.[9] Conversely, if the correlations are lower, the achieved statistical power would be lower.

**Table 17**

*Sample characteristics in the two experiments*

| Sample Number | Data Collection Month | Sample Platform | *N* | $M_{age}$ | $SD_{age}$ | % Men | % Women | % White | % Black | % Other Ethnic |
|---|---|---|---|---|---|---|---|---|---|---|
| Exp 1 | December 2024 | Prolific | 2789 | 38.42 | 13.09 | 49.08 | 48.26 | 58.87 | 25.89 | 15.24 |
| Exp 2 | February 2025 | Prolific | 1564 | 39.04 | 13.32 | 48.59 | 48.21 | 70.01 | 9.84 | 20.15 |

**Materials and Procedure**

***Independent Variable***

Participants were randomly assigned to one of six conditions: five moral humility interventions or treatments or a control (about artificial intelligence). Interventions are described in SOM.

***Dependent Variables***

A summary of the dependent variables included in the experiment is in Table 12.

**Moral Humility.** The main outcome measure was moral humility. This was measured on a 9-item moral humility scale ($\alpha = 0.82$), a shorter moral humility scale that included a subset of the 30-item scale constructed and used in the previous studies. 3-items each were sampled from each of the moral humility subscales based on genetic algorithm used to construct shorter scales (Schroeder et al., 2016). The items used are indicated in Table 3.

[9] This power calculation is different from what was preregistered, as there was a mistake made in the preregistered power calculation. The preregistered power analysis underestimated the power of the sample.

The impact of the interventions was also explored at the subscale level, i.e., with moral learning/openness (α = 0.85), moral fallibility (α = 0.83), and moral superiority (α = 0.76). This 9-item shorter scale was used instead of the full scale to keep the study a manageable length.

**Political Outcomes.** Three political outcomes were measured. *Political sectarianism* (Finkel et al., 2024) was assessed as it is a measure meant to capture the moralized nature of partisan disdain in the US context. The 9-item measure of political sectarianism (Finkel et al., 2024) was used (α = 0.97), where participants indicated their agreement (*1 = strongly disagree, 7 = strongly agree*) on the scale's following items which were averaged to create a measure: "I am different from the typical [Republican/Democrat].", "I feel distant from the typical [Republican/Democrat]", "No matter how hard I try, I can't see the world the way the typical [Republican/Democrat] does.", "I hate the typical [Republican/Democrat].", "My feelings toward the typical [Republican/Democrat] are negative", "The typical [Republican/Democrat] has lots of negative traits.", "The typical [Republican/Democrat] is immoral.", "The typical [Republican/Democrat] is evil.", "The typical [Republican/Democrat] lacks integrity." The scale included three subscales capturing othering (first three items; α = 0.93), aversion (middle three items; α = 0.91), and moralization (last three items; α = 0.94). Accordingly, I explored the effect of the interventions on the subscale separately as well, in addition to testing the intervention's effect on the whole scale.

*Selective exposure* was assessed in the same way as in Samples 4 and 5 in Chapter II, wherein participants indicated how interested they were in hearing from someone who held the opposing view on an important issue (*-100 = very uninterested, 0 = neutral, 100 = very interested*). However, for the sake of keeping the study short, participants completed the selective exposure task for only one issue that they selected was an important issue for them. *Political compromise* was measured the same way as in Sample 6 in Chapter II, wherein participants indicated how likely they were to vote for a compromising candidate and an uncompromising candidate (*1 = not at all, 7 = very likely)* who differed in their approach to negotiating on an issue important to the participant. The main difference in both selective exposure and political compromise tasks compared to the prior studies was in the assignment of issues for each task, i.e., on which issues the participants completed the task and how it is chosen for each participant. In Experiment 1, participants indicated the two important issues for them from a list of issues. These issues were the eleven issues used in Sample 4 and 5. One of these issues was used for the selective exposure

task and one for the political compromise task. The importance of the issue and participant's position on the issue for selective exposure were both assessed *before* participants underwent the moral humility intervention.

**Moral Relativism and Intellectual Humility**. Moral relativism and intellectual humility, two constructs that have had strong correlations with moral humility in previous studies were also measured to see if and the extent to which moral humility interventions have an impact on them. Moral relativism was measured using a subset of 10-item scale (Collier-Spruel et al., 2019). Specifically, six highest loading items were chosen while balancing the breadth of content covered ($\alpha = 0.71$). Example items were: "There is a moral standard that all actions should be held to, even if cultures disagree.", "There are moral rules that apply to everyone regardless of personal beliefs.". See SOM for full list of items. Again, fewer items were used to keep the study a manageable length for participants. Intellectual humility was measured using the same 6-item (Leary et al., 2017) scale used in all the previous studies ($\alpha = 0.89$).

**Emotions.** As in the Pilot, participants reported the intensity of various epistemic emotions they experienced during the reading of the interventions on the Epistemically-Related Emotion Scales (Pekrun et al., 2017). Only the six most informative and simple ones, i.e., surprised, curious, confused, anxious, frustrated, and bored were included in Experiment 1 to keep the study a manageable length. Participants answered these on a 7-point scale *(1=Not at All, 4=Moderate, 7=Very Strong).* Again, the aim was to see if the interventions impacted these emotions differently, which would be used to provide insight into any possible differences between the effectiveness of these interventions.

***Pre-Treatment Covariates***

The experiment used a pre-post experimental design; a pre-post experimental design involves measuring the dependent or outcome variable both before and after the treatment. This type of design has been shown to increase precision of estimates as well as provide more power to detect an effect (Clifford et al., 2021), given that the pre- and post- measures are highly correlated. The design essentially allows the examination of how participants' attitudes change over the course of a study and whether the pattern of this change differs among those assigned to different conditions in an experiment.

Accordingly, outcome variables were measured before the treatment as well, which were then treated as covariates in the analysis (Clifford et al., 2021). Specifically, in Experiment 1, moral humility (Moral Humility $\alpha = 0.78$; Moral Learning/Openness $\alpha = 0.81$; Moral Fallibility $\alpha = 0.78$; Moral Superiority $\alpha = 0.71$) and political

sectarianism (Political Sectarianism α = 0.97; Othering α = 0.91; Aversion α = 0.90; Moralization α = 0.93) were measured before the treatment (using the same post-treatment measures described above). Pre-treatment measures of selective exposure and political compromise were not included as I conjectured that pre-treatment political sectarianism might be highly correlated with all three political outcomes (political sectarianism, selective exposure, political compromise) and serve as a useful pre-treatment covariate for all. Similarly, pre-treatment moral humility was considered a pre-treatment measure for intellectual humility and moral relativism also, apart from post-treatment moral humility. These decisions were made to keep the study a manageable length. If all outcomes were measured before the treatment, that would have made the study very long. A summary of pre-treatment variables included in the experiment is provided in Table 12.

***Exploratory Variables***

**Open-Ended Report.** Participants also reported their general thoughts in an open-ended box at the end of the treatments or control. They read: "You read about (moral) ideas that philosophers and psychologists have thought about, written about, and studied. Please take a few moments to think about the things you read. Please tell us very briefly: What did you think about the ideas you just read about? Was there any information that challenged or expanded your thinking (about morality)?". This was collected for two purposes. Primarily, it was collected under the guise of the main outcome variable we are interested in as researchers. The whole task or experiment was set up such that participants were told that we researchers are interested in what they think of various philosophical views. After reading the various vignettes, this open-ended question would serve as the place where researchers collect the participant's views. Secondarily, these responses were intended to be used as a potential exploratory variable which would be qualitatively analyzed to understand how the participants perceived the vignettes. This was however not a variable of interest and was more of a filler question as part of the cover story and was intended to be analyzed only if time permits.

**Analysis and Results**

The aim of the study was to examine if the five moral humility interventions or treatments each increase moral humility and decrease the polarization outcomes compared to the control condition, and whether the effect of the treatments on the polarization outcomes is mediated by moral humility. The effects of the five treatments on intellectual humility, moral relativism, and epistemic emotions was also examined. All preregistered analyses and predictions can be found at OSF.

***Moral Humility***

To test the interventions' effect on moral humility, linear regression was used wherein post-treatment moral humility was regressed on five dummy-coded condition variables and pre-treatment levels of moral humility. Each dummy code compared each of the five treatments to the control. The prediction was that people assigned to the five moral humility intervention conditions will show higher moral humility compared to those in the control condition. The correlation between pre- and post- moral humility was $r = 0.82$, $p < .001$.

Consistent with the predictions, moral humility was significantly higher in each of the five experimental or treatment conditions compared to the control condition (Table 18). Results are shown in Figure 12. The average effect size or Cohen's *d* across the five treatments was $d = 0.23$ [0.14, 0.33].

**Figure 12**

*Estimated difference in moral humility between the treatments and control (accounting for pre-treatment levels of moral humility)*


*Note:* X-axis coefficients (*b*) are on the original 1-7 scale.

**Table 18**

*Estimates (and standard errors) of the difference in moral humility between each of the five treatments versus control. These are estimates for Figure 12*

| | Moral Humility<br>*b* | Effect Size<br>Cohen's *d* |
|---|---|---|
| Biased Worldviews | 0.18** <br>(0.04) | 0.33 |
| Difficulty Ethics | 0.12** <br>(0.04) | 0.22 |
| Evil Capacity | 0.08* <br>(0.04) | 0.14 |
| Imperfect Past | 0.14** <br>(0.04) | 0.21 |
| Moral Disagreements | 0.12** <br>(0.04) | 0.22 |
| Moral Humility<br>(Pre-Treatment) | 0.89** <br>(0.01) | - |
| Observations | 2,755 | |

*Note:* *p<0.05; **p<0.01. Regression coefficients (*b*) are on the original 1-7 scale. The effect size or Cohen's *d* was computed by dividing the marginal means by residual standard deviation in an ANOVA.

The effect of each of the five treatments on the three moral humility subscales was also examined. The results for the subscales showed that different aspects of moral humility (moral learning/openness, moral fallibility, moral superiority) contributed to the treatment effect of the five interventions to different extent. For example, biased worldview and imperfect past interventions most strongly impacted moral fallibility, moral disagreement intervention most strongly impacted moral learning/openness, and the difficulty ethics and evil capacity interventions most strongly impacted moral superiority. Results are in Table 19.

Moderation analysis was not preregistered but explored. Specifically, I examined if the effect of the five interventions on moral humility was moderated by party identity or education to assess if the interventions work similarly across people of different political leanings and educational levels. To test this, a party identity dummy variable (Democrat/Democrat leaning = 0, Republican/Republican leaning = 1) and education dummy variable (less than bachelor's education = 0, bachelor's education and higher = 1) was created and an interaction variable between these and the experimental condition dummy variable was added in the regression. There weren't any significant interactions, suggesting that the interventions worked similarly across political leanings and education levels.

Finally, I also examined if the five interventions significantly differed from each other using post-hoc pairwise comparisons in ANOVA. Across ten pairwise comparison, there were no significant differences between the interventions.

**Table 19**

*Estimates (and standard errors) of the difference in moral humility subscales between each of the five treatments versus control*

| | Moral Learning<br>*b* | Moral Fallibility<br>*b* | Moral Superiority<br>*b* |
|---|---|---|---|
| Biased Worldviews | 0.14$^{**}$<br>(0.05) | 0.26$^{**}$<br>(0.06) | -0.09$^{\dagger}$<br>(0.06) |
| Difficulty Ethics | 0.08$^{\dagger}$<br>(0.05) | 0.12$^{\dagger}$<br>(0.06) | -0.15$^{**}$<br>(0.06) |
| Evil Capacity | 0.05<br>(0.05) | 0.08<br>(0.06) | -0.10$^{\dagger}$<br>(0.06) |
| Imperfect Past | 0.08$^{\dagger}$<br>(0.05) | 0.20$^{**}$<br>(0.06) | -0.07<br>(0.06) |
| Moral Disagreements | 0.12$^{**}$<br>(0.05) | 0.10$^{\dagger}$<br>(0.06) | -0.09<br>(0.06) |
| Moral Learning<br>(Pre-Treatment) | 0.84$^{**}$<br>(0.01) | | |
| Moral Fallibility<br>(Pre-Treatment) | | 0.77$^{**}$<br>(0.01) | |
| Moral Superiority<br>(Pre-Treatment) | | | 0.76$^{**}$<br>(0.01) |
| Observations | 2,693 | 2,694 | 2,694 |

*Note:* $^{\dagger}p\leq0.1$, $^{*}p\leq0.05$; $^{**}p\leq0.01$. Regression coefficients (*b*) are on the original 1-7 scale. Marginally significant results are also highlighted as analysis at subscale level can reduce power and precision.

Taken together, the results suggested that all five moral humility interventions significantly increased moral humility compared to the control, the five interventions were not significantly different from each other. The interventions worked similarly in increasing moral humility for Democrats and Republicans and people of different education levels, with a variation in how strong the effects of the interventions were on the outcome.

***Political Outcomes***

To test the interventions' effect on polarization, three political outcomes — political sectarianism, selective exposure, political compromise — were examined, and a similar test was conducted as moral humility, but on the political outcomes. That is, linear regression was used to test if the three political variables significantly differed between each of five treatments and control. For each political outcome analysis, I controlled for pre-treatment levels of political sectarianism. The prediction was that compared to the control condition, people in the moral

humility intervention conditions will express (i) lower political sectarianism, (ii) more interest in cross-cutting exposure or exposure to opposing political viewpoint, and (iii) more willingness towards political compromise.

Notably, a logic error made in the study's Qualtrics program led to half the sample getting the wrong the political outgroup for the political sectarianism measure. This compromised the power for the analyses of the three political outcomes as approximately half the sample had to be excluded for the preregistered analyses for these political outcomes. Regardless of this mistake, the experiment was able to still provide good, high-powered (80% power to detect $d = 0.15$) information for at least one of these outcomes, i.e., political sectarianism. This was because the pre-treatment political sectarianism measure was highly correlated with the post-treatment political sectarianism measure ($r = 0.90$, $p < .001$), which helped enhance power. However, the tests for the other two political outcomes, i.e., selective exposure and political compromise, didn't have good power (~ 40 - 55% power to detect $d = 0.2$) because of the loss of sample and low-correlation with pre-treatment political sectarianism ($r \sim 0.15$ - 0.30). Given these, I am more confident in the political sectarianism results than those for selective exposure and political compromise.

**Political Sectarianism.** Political sectarianism was significantly higher for three of the five treatments compared to the control. These treatments were biased worldviews, evil capacity, and moral disagreements (Table 20). Results are shown in Figure 13. The average effect size or Cohen's $d$ across the five treatments was $d = 0.21$ [0.11, 0.30].

The effect of each of the five interventions on the three political sectarianism subscales (othering, aversion, moralization) was also examined. The results for the subscales show that different aspects of political sectarianism (othering, aversion, moralization) contribute to the treatment effect of the five interventions to different extent (Table 21). All five interventions had the biggest impact on the othering aspect of political sectarianism, followed largely by aversion, and then moralization.

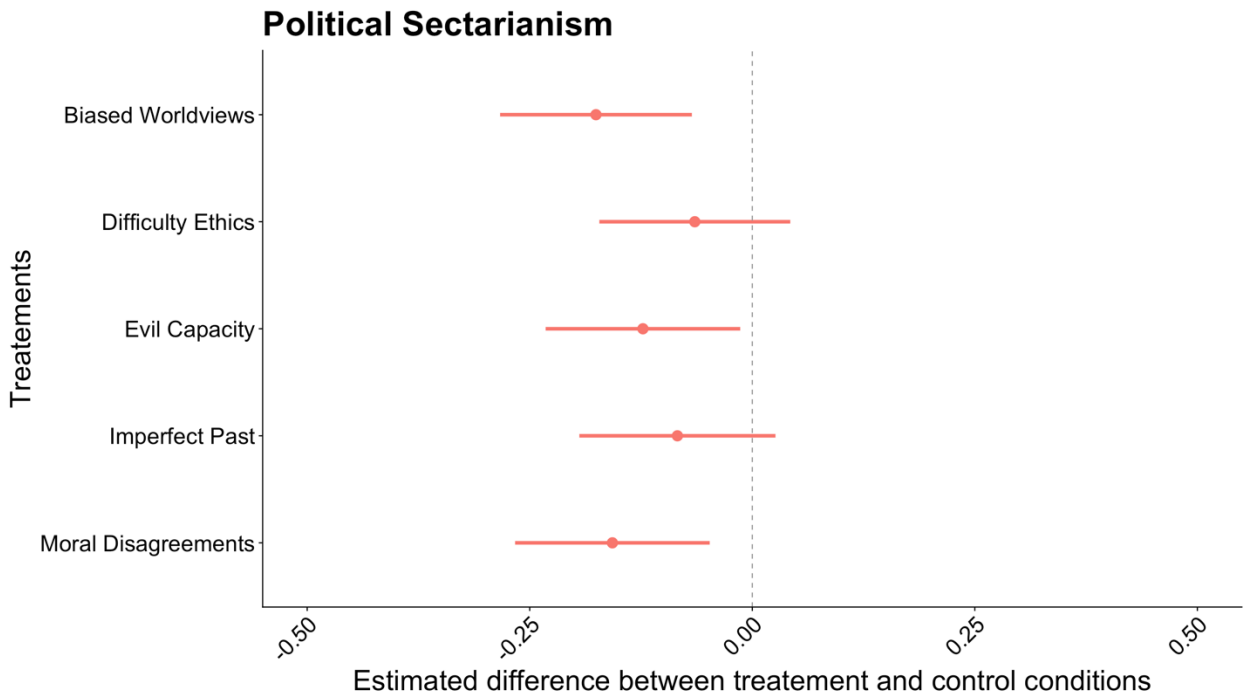
Like with moral humility outcome, I conducted moderation analysis (not preregistered, was exploratory) to examine if the effect of the five interventions on political sectarianism was moderated by party identity or education. Largely, there weren't any significant interactions of both with the five interventions, suggesting that the interventions worked similarly in reducing political sectarianism across political leanings and education levels. There was one marginally significant interaction — between imperfect past intervention and party identity ($b = 0.21$,

$p = 0.059$) — suggesting that it might be that imperfect past intervention is more effective for reducing political sectarianism for those who identify as Democrats than Republicans.

I also examined if the five interventions significantly differed from each other in reducing political sectarianism using post-hoc pairwise comparisons in ANOVA. Across ten pairwise comparison, I found no evidence of significant differences between the interventions.

**Figure 13**

*Estimated difference in political sectarianism between the treatments and control (accounting for pre-treatment levels political sectarianism)*




*Note:* X-axis coefficients (*b*) are on the original 1-7 scale.

**Table 20**

*Estimates (and standard errors) of the difference in political sectarianism between each of the five treatments versus control. These are estimates for Figure 13*

| | Political Sectarianism *b* | Effect Size Cohen's *d* |
|---|---|---|
| Biased Worldviews | -0.18** (0.06) | 0.30 |
| Difficulty Ethics | -0.07 (0.06) | 0.11 |
| Evil Capacity | -0.12* (0.06) | 0.21 |
| Imperfect Past | -0.09 (0.06) | 0.14 |
| Moral Disagreements | -0.16** (0.06) | 0.27 |
| Political Sectarianism (Pre-Treatment) | 0.96** (0.06) | - |
| Observations | 1369 | |

*Note:* $^{*}p \leq 0.05$; $^{**}p \leq 0.01$. Regression coefficients (*b*) are on the original 1-7 scale.

**Table 21**

*Estimates (and standard errors) of the difference in political sectarianism subscales between each of the five treatments versus control*

| | Political Sectarianism Othering *b* | Political Sectarianism Aversion *b* | Political Sectarianism Moralization *b* |
|---|---|---|---|
| Biased Worldviews | -0.22** (0.06) | -0.19** (0.07) | -0.13† (0.07) |
| Difficulty Ethics | -0.12 (0.06) | -0.08 (0.07) | -0.01 (0.07) |
| Evil Capacity | -0.13† (0.06) | -0.13† (0.07) | -0.11 (0.07) |
| Imperfect Past | -0.16* (0.06) | -0.05 (0.07) | -0.05 (0.07) |
| Moral Disagreements | -0.19* (0.06) | -0.16* (0.07) | -0.12 (0.07) |
| Political Sectarianism Othering (Pre-Treatment) | 0.93** (0.01) | | |
| Political Sectarianism Aversion (Pre-Treatment) | | 0.93** (0.011) | |
| Political Sectarianism Moralization (Pre-Treatment) | | | 0.93** (0.012) |
| Observations | | 1,369 | |

$^{\dagger}p \leq 0.1$, $^{*}p \leq 0.05$; $^{**}p \leq 0.01$. Regression coefficients (*b*) are on the original 1-7 scale. Marginally significant results are also highlighted as analysis at subscale level can reduce power and precision.

***Political Sectarianism Mediation.*** The results of the mediation analysis are in Table 22, wherein I tested the extent to which moral humility explained the effects of the interventions on political sectarianism. The ACME (Average Causal Mediation Effect) tells us how much of the effect of the interventions on political sectarianism is mediated by moral humility. Thus, it is the indirect effect of the interventions on political sectarianism through moral humility. If ACME is significant (i.e., confidence interval doesn't contain zero)**,** it suggests that moral humility as the mediator plays a significant role in explaining the relationship between the interventions (moral humility treatments) and political sectarianism (outcome).

The ADE (Average Direct Effect) is part of effect of the interventions on political sectarianism that is not mediated by moral humility. Thus, it is the direct effect of the interventions on political sectarianism after accounting for moral humility. If ADE is significant, it means the interventions still influence political sectarianism even after controlling for moral humility. The Total Effect is the overall effect of the interventions on political sectarianism (ACME + ADE). The Proportion Mediated tells us how much of the Total Effect is explained moral humility.

The expectation was that the ACME's will be negative and significant, indicating that the interventions increase moral humility and an increase in moral humility leads to decrease in political sectarianism. Result showed that moral humility significantly mediated (ACME) the effect of the four interventions on political sectarianism. These were: biased worldviews, imperfect past, evil capacity, and moral disagreements. The interventions also had significant direct and total effects on political sectarianism for two of these interventions, i.e., biased worldviews and moral disagreements.

**Table 22**

*Mediation results for the effect of the interventions on political sectarianism via moral humility*

| | Difficulty Ethics | Biased Worldviews | Imperfect Past | Evil Capacity | Moral Disagreements |
|---|---|---|---|---|---|
| ACME | -0.01 | -0.04** | -0.03** | -0.04** | -0.03** |
| ADE | -0.05 | -0.12* | -0.02 | -0.08 | -0.12* |
| Total Effect | -0.06 | -0.16** | -0.05 | -0.12* | -0.15** |
| Prop Mediated | 0.19 | 0.27 | 0.66 | 0.33 | 0.20 |

*Note:* $^{*}p \leq 0.05$; $^{**}p \leq 0.01$.

I also conducted sensitivity analysis to address the sequential ignorability assumption (Imai et al., 2010; Tingley et al., 2014). To that end, the correlation between the residuals of the mediator and outcome regressions was

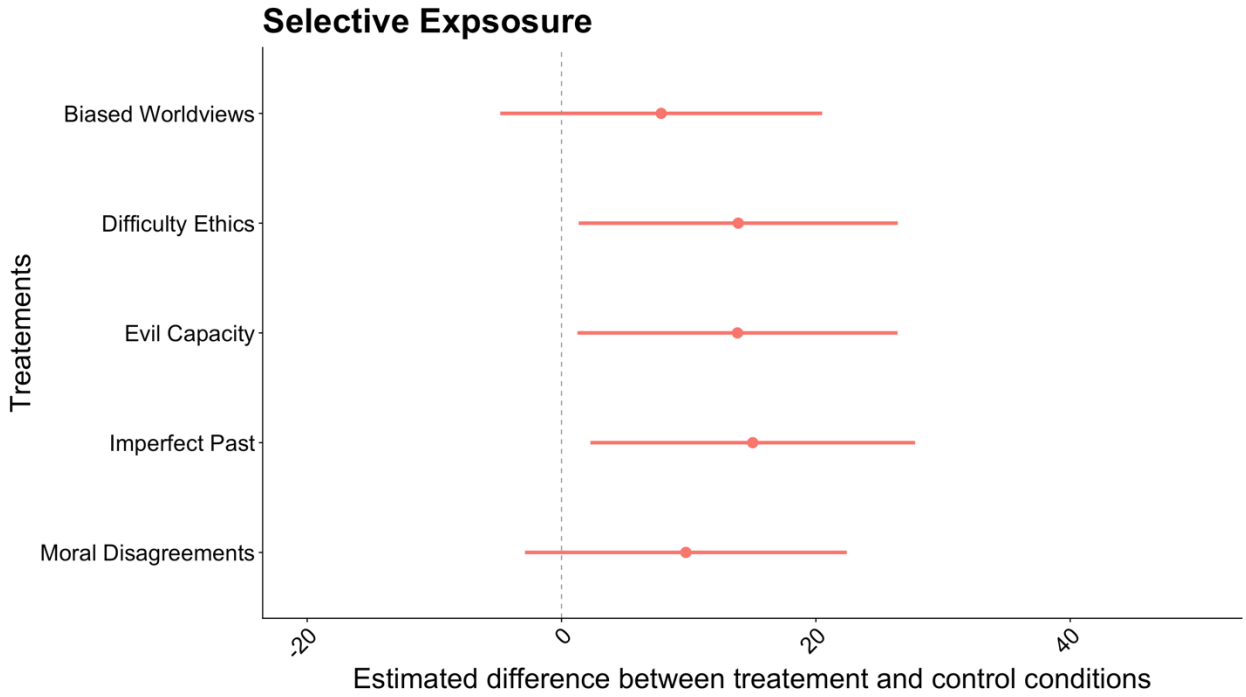
chosen as the sensitivity parameter. This is because the relationship between moral humility and political outcomes was not causal, and thus there existed the possibility of the existence of unobserved pre-treatment confounders affecting both the mediator and the outcome, making the correlation between the residuals not zero. The sensitivity analysis varied the value of this correlation between -0.9 and +0.9 by 0.1 increments and examined how the estimated indirect effect changes. The results of this sensitivity analysis showed that across the significant mediation models, a correlation of $r \sim 0.1$- 0.2 due to unmeasured confounders could nullify the mediation effect.

Taken together, the results suggested that only three of the five moral humility interventions — biased worldviews, evil capacity, and moral disagreements — significantly reduced political sectarianism compared to the control. The effects of all three was mediated by moral humility. However, the five interventions were not significantly different from each other. Two of these might have been not significantly different from control due to weaker treatment effect (e.g., difficulty ethics) or heterogenous effects across groups weakening the overall treatment effect (e.g., imperfect past); in fact, for one of these (i.e., imperfect past), there was a significant indirect effect through moral humility. However, sensitivity analysis suggests that the mediation models in general are very sensitive to the existence of confounders. Overall, the results suggested that the interventions worked similarly in reducing political sectarianism, had similar effects for Democrats and Republicans and people of different education levels, with a variation in how strong the effects of the five interventions were on the outcome.

**Selective Exposure.** The willingness towards cross-cutting exposure or engaging with opposing political viewpoint on the participant's self-selected important issue was significantly higher for three of the five treatments compared to the control. These treatments were difficulty ethics, evil capacity, and imperfect past (Table 23). Results are shown in Figure 14. The average effect size or Cohen's *d* across the five treatments was $d = 0.2$ [0.13, 0.25].

**Figure 14**

*Estimated difference in selective exposure between the treatments and control (accounting for pre-treatment levels of political sectarianism)*




*Note:* X-axis coefficients (*b*) are on the original -100 - +100 scale. Higher and positive values indicate more interest in exposure to opposing viewpoint.

**Table 23**

*Estimates (and standard errors) of the difference in selective exposure between each of the five treatments versus control. These are estimates for Figure 14*

| | Selective Exposure<br>*b* | Effect Size<br>Cohen's *d* |
|---|---|---|
| Biased Worldviews | 7.84<br>(6.46) | 0.13 |
| Difficulty Ethics | 13.90*<br>(6.40) | 0.23 |
| Evil Capacity | 13.84*<br>(6.42) | 0.23 |
| Imperfect Past | 15.04*<br>(6.51) | 0.25 |
| Moral Disagreements | 9.78<br>(6.45) | 0.16 |
| Political Sectarianism<br>(Pre-Treatment) | -13.06**<br>(1.142) | - |
| Observations | 1,058 | |

Table 23 (cont'd)
*Note*: $^{*}p \leq 0.05$; $^{**}p \leq 0.01$. Regression coefficients (*b*) are on the original -100 - +100 scale.

Like before, I conducted moderation analysis (not preregistered, was exploratory) to examine if the effect of the five interventions on selective exposure was moderated by party identity or education. Largely, there weren't any significant interactions of both with the five interventions, suggesting that the interventions worked similarly in reducing political sectarianism across political leanings and education levels. There was one significant interaction — between imperfect past intervention and education level ($b = 27.47$, $p = 0.043$) — suggesting that it might be that imperfect past intervention is more effective for increasing interest in cross cutting exposure for those with bachelor's education or higher.

I also examined if the five interventions significantly differed from each other in increasing cross-cutting exposure using post-hoc pairwise comparisons in ANOVA. Across ten pairwise comparison, I found no evidence of significant differences between the interventions.

***Selective Exposure Mediation.*** The results of the mediation analysis are in Table 24, wherein I tested the extent to which moral humility explained the effects of the interventions on interest in cross-cutting exposure. The expectation was that the ACME's will be positive and significant, indicating that the interventions increase moral humility and an increase in moral humility leads to increase in interest in cross-cutting exposure.

Results showed that moral humility mediated (ACME) the effect of the four interventions on willingness to engage with opposing political viewpoint either significantly or marginally significantly. These were: difficulty ethics, biased worldviews, imperfect past, and evil capacity. There were two (difficulty ethics, evil capacity) and three (difficulty ethics, evil capacity, moral disagreements) marginally significant or significant direct and total effects respectively.

**Table 24**

*Mediation results for the effect of the interventions on selective exposure via moral humility*

| | Difficulty Ethics | Biased Worldviews | Imperfect Past | Evil Capacity | Moral Disagreements |
|---|---|---|---|---|---|
| ACME | 2.05$^{\dagger}$ | 4.65$^{**}$ | 4.36$^{*}$ | 2.42$^{\dagger}$ | 1.80 |
| ADE | 12.23$^{*}$ | 0.41 | 6.45 | 9.89$^{\dagger}$ | 8.08 |
| Total Effect | 14.41$^{*}$ | 5.05 | 10.82 | 12.31$^{*}$ | 9.88$^{\dagger}$ |
| Prop Mediated | 0.15 | 0.91 | 0.40 | 0.20 | 0.18 |

*Note*: $^{\dagger}p \leq 0.1$, $^{*}p \leq 0.05$; $^{**}p \leq 0.01$. Marginally significant results are also highlighted as analysis had lower power due to exclusion of sample that got wrong political items.

I also conducted sensitivity analysis, following the same procedure as used before for political sectarianism mediation models. Like before, the results of the sensitivity analysis showed that across the mediation models, a correlation of $r \sim 0.1$- 0.2 due to unmeasured confounders could nullify the mediation effect.

Taken together, the results suggested that only three of the five moral humility interventions — difficulty ethics, evil capacity, and imperfect past — significantly increase cross-cutting exposure compared to the control. The effects of all three was mediated by moral humility. However, the five interventions were not significantly different from each other. Two of these might have been not significantly different from control due to weaker treatment effects; in fact, for one of these (i.e., biased worldviews), there was a significant indirect effect through moral humility. However, sensitivity analysis suggests that the mediation models in general are sensitive to the existence of confounders. Overall, the results suggested that the interventions worked similarly in increasing cross-cutting exposure, had similar effects for Democrats and Republicans and people of different education levels, with a variation in how strong the effects of the five interventions were on the outcome. However, all of these results should be interpreted tentatively until they are replicated, given the lower power (~ 60%) for the selective exposure analyses.

**Political Compromise.** The five treatments did not significantly impact support towards uncompromising candidate (candidate A), the compromising candidate (candidate B), or compromising over uncompromising candidate (candidate B-A) (Table 25) compared to the control. The average effect size or Cohen's *d* across the five treatments for support towards candidate A was $d = 0.08$ [0.01, 0.16], towards candidate B was $d = -0.08$ [-0.18, 0.05], and towards candidate B v A was $d = 0.09$ [-0.18, 0.02].

**Table 25**

*Estimates (and standard errors) of the difference in support for candidate A (uncompromising candidate), candidate B (compromising candidate), and candidate B vs A (uncompromising over compromising candidate), between each of the five treatments versus control*

| | Candidate A *b* | Candidate A *d* | Candidate B *b* | Candidate B *d* | Candidate B-A *b* | Candidate B-A *d* |
|---|---|---|---|---|---|---|
| Biased Worldviews | 0.25 (0.18) | 0.13 | -0.12 (0.14) | -0.08 | -0.36 (0.28) | -0.12 |
| Difficulty Ethics | 0.16 (0.18) | 0.08 | -0.07 (0.14) | -0.05 | -0.24 (0.28) | -0.08 |
| Evil Capacity | 0.30 (0.18) | 0.16 | -0.27$^{\dagger}$ (0.14) | -0.18 | -0.53$^{\dagger}$ (0.28) | -0.18 |
| Imperfect Past | 0.02 (0.18) | 0.01 | 0.07 (0.14) | 0.05 | 0.05 (0.28) | 0.02 |
| Moral Disagreements | 0.12 (0.18) | 0.06 | -0.21 (0.14) | -0.15 | -0.31 (0.28) | -0.11 |
| Political Sect | 0.18$^{**}$ (0.03) | - | -0.14$^{**}$ (0.02) | - | -0.32$^{**}$ (0.05) | - |
| Observations | 1,355 | | 1,367 | | 1,353 | |

*Note*: $^{\dagger}p\leq0.1$, $^{*}p\leq0.05$; $^{**}p\leq0.01$. Regression coefficients (*b*) are on the original 1-7 scale. Marginally significant results are also highlighted as analysis had lower power due to exclusion of sample that got wrong political items.

As before, I conducted moderation analysis (not preregistered, was exploratory) to examine if the effects of the five interventions on the three political compromise measures were moderated by party identity or education. There weren't any significant interactions of party identity or education with the five interventions for all three political compromise measures, suggesting that the interventions worked similarly for political compromise across diverse political leanings and education levels.

I also examined if the five interventions significantly differed from each other in increasing political compromise across the three measures using post-hoc pairwise comparisons in ANOVA. Across ten pairwise comparisons, I found no evidence of significant differences between the interventions.

***Political Compromise Mediation*.** The results of the mediation analysis are in Table 26, wherein the extent to which moral humility explained the effects of the interventions on support for uncompromising (candidate A),

compromising (candidate B), and compromising over uncompromising (candidate B-A) candidate was examined. The expectation was that the ACME for support for the uncompromising (candidate A) will be negative and significant, indicating that the interventions increase moral humility and an increase in moral humility leads to decrease in support for uncompromising (candidate A). Next, for support for the compromising (candidate B) and support for the compromising over uncompromising candidate (candidate B-A), the expectation was that ACME's will be positive and significant, indicating that the interventions will increase moral humility and an increase in moral humility leads to an increase in support for compromising (candidate B) and support for compromising over uncompromising candidate (candidate B-A).

Results showed that the ACME's tended to be significantly or marginally significant in the direction that was expected for almost all five interventions (Table 26). The interventions increased moral humility and consequently — decreased support for uncompromising candidate, increased support for the compromising candidate, and increased support for the compromising over uncompromising candidate. However, the direct and total effects (whenever significant or marginally significant) were in the opposite direction to indirect effects. This pattern of direct and indirect results being in the opposing directions could be because it is possible that while the interventions do increase willingness towards political compromise via increased moral humility, they also decrease willingness towards political compromise via another mechanism. In other words, there might be competing mediators that counteract the positive effects of moral humility, resulting in overall negative effect.

Further, I also conducted sensitivity analysis, following the same procedure as used before for political sectarianism and selective exposure mediation models. Like before, the results of the sensitivity analysis showed that across the mediation models, a correlation of $r \sim 0.1$- 0.2 due to unmeasured confounders could nullify the mediation effect.

In any case, all of these results should be interpreted tentatively, given the non-significant effects of the interventions on political compromise outcomes (Table 25) and the pretty low power (~40%) for these analyses.

**Table 26**

*Mediation results for the effect of the interventions on support for candidate A (uncompromising candidate), candidate B (compromising candidate), candidate B vs A (uncompromising over compromising candidate) via moral humility*

| | Difficulty Ethics | Biased Worldviews | Imperfect Past | Evil Capacity | Moral Disagreements |
|---|---|---|---|---|---|
| | | | Candidate A (Uncompromising) | | |
| ACME | -0.05 | -0.11** | -0.08* | -0.05† | -0.12** |
| ADE | 0.2 | 0.37* | 0.17 | 0.37* | 0.25 |
| Total Effect | 0.16 | 0.27 | 0.09 | 0.31† | 0.13 |
| Prop Mediated | -0.31 | -0.39 | -0.82 | -0.18 | -0.87 |
| | | | Candidate B (Compromising) | | |
| ACME | 0.04* | 0.09* | 0.10** | 0.09** | 0.11** |
| ADE | -0.13 | -0.24† | -0.13 | -0.38** | -0.34** |
| Total Effect | -0.09 | -0.16 | -0.03 | -0.29* | -0.23 |
| Prop Mediated | -0.49 | -0.57 | -3.05 | -0.3 | -0.50 |
| | | | Candidate B-A (Compromising v Uncompromising) | | |
| ACME | 0.09* | 0.2** | 0.18** | 0.15** | 0.23** |
| ADE | -0.34 | -0.62* | -0.30 | -0.72** | -0.58* |
| Total Effect | -0.25 | -0.42 | -0.12 | -0.57* | -0.35 |
| Prop Mediated | -0.37 | -0.47 | -1.51 | -0.27 | -0.69 |

*Note*: †$p \leq 0.1$, *$p \leq 0.05$; **$p \leq 0.01$. Marginally significant results are also highlighted as analysis had lower power due to exclusion of sample that got wrong political items.

Taken together, the results suggested that the interventions worked similarly, i.e., didn't reliably impact political compromise across all political compromise outcomes, and had similar effects across political leanings and education levels. There was some suggestive evidence that the interventions might indeed be increasing willingness towards political compromise via increased moral humility, but that there are also other mechanisms at play which might be reducing or counteracting the positive effects of moral humility on political compromise. These results would need to be replicated in future studies with good power before any of these conclusions can be made confidently or warrant any further investigation or speculation.

***Intellectual Humility and Moral Relativism***

The analysis for moral relativism and intellectual humility followed the same analytic strategy as that for moral humility. The correlation between pre-treatment moral humility and post-treatment intellectual humility was $r = 0.56$, $p < .001$, and with post-treatment moral relativism was $r = 0.47$, $p < .001$.

All five interventions did not significantly in(de)crease intellectual humility or moral relativism (Table 27). The average effect size or Cohen's *d* across the five treatments for intellectual humility was *d* = -0.01 [-0.05, 0.01], and for moral relativism was *d* = -0.02 [-0.05, 0.07]. Again, these effects were not moderated by party identity or educational levels.

Taken together, these results (compared to moral humility results which increased across all five treatments) provide evidence for moral humility's discriminant validity with both these psychological constructs.

**Table 27**

*Estimates (and standard errors) of the difference in intellectual humility and moral relativism between each of the five treatments versus control*

| | Intellectual Humility *b* | Intellectual Humility *d* | Moral Relativism *b* | Moral Relativism *d* |
|---|---|---|---|---|
| Biased Worldviews | 0.01 (0.06) | 0.006 | -0.05 (0.06) | -0.05 |
| Difficulty Ethics | -0.05 (0.06) | -0.04 | 0.02 (0.06) | 0.03 |
| Evil Capacity | -0.05 (0.06) | -0.04 | -0.02 (0.06) | -0.02 |
| Imperfect Past | 0.01 (0.06) | 0.01 | -0.06 (0.06) | -0.07 |
| Moral Disagreements | 0.01 (0.06) | 0.01 | 0.07 (0.06) | 0.08 |
| Moral Humility (pre-treatment) | 0.63** (0.02) | - | 0.53** (0.02) | - |
| Observations | 2,692 | | 2,691 | |

*Note*: *$p \leq 0.05$; **$p \leq 0.01$

***Epistemic Emotions***

Six epistemic emotions — surprise, curiosity, confusion, anxiety, frustration, and boredom — were measured in all five experimental and control conditions. This was to assess if the interventions differed in the emotions they evoked, which would help understand any differences in the impact of these interventions on moral humility or political outcomes. All six conditions were compared pairwise using ANOVA (adjusting for Tukey's multiple comparisons). The results are in Table 28.

Results indicated that the moral disagreement treatment made people significantly less surprised and curious compared to the biased worldview and imperfect past treatments. There were no differences in confusion and boredom across all pairwise comparisons. Compared to the control, evil capacity and imperfect past vignettes evoked significantly more frustration. Evil capacity treatment evoked significantly more anxiety and frustration than

the biased worldviews and moral disagreement treatments. Evil capacity also evoked more frustration compared to the difficulty ethics treatment.

Overall, there was largely a lack of very consistent patterns in these comparisons, barring one. Evil capacity treatment appeared to have the most impact on the negative emotions, especially anxiety and frustration. This might help explain why that treatment was the one that had the smallest impact on moral humility, although as we saw before, the impact was not significantly different than the effect of other treatments. Thus, at this point, it was unclear what the higher levels of frustration and anxiety in the evil capacity treatment might mean substantively for its effect on outcomes.

**Table 28**

*Pairwise comparisons of epistemic emotions between the six (five treatments and one control) conditions*

| Contrast | Surprised | Curious | Confused | Anxious | Frustrated | Bored |
|---|---|---|---|---|---|---|
| Artificial Intelligence - | -0.05 | 0.01 | -0.03 | 0.10 | -0.14 | -0.13 |
| Biased Worldviews | (0.12) | (0.11) | (0.09) | (0.10) | (0.10) | (0.09) |
| Artificial Intelligence - | 0.11 | 0.13 | -0.10 | -0.04 | -0.11 | -0.15 |
| Difficulty Ethics | (0.12) | (0.11) | (0.09) | (0.10) | (0.10) | (0.09) |
| **Artificial Intelligence -** | 0.07 | 0.25 | -0.08 | -0.28 | **-0.53** | 0.01 |
| **Evil Capacity** | (0.12) | (0.11) | (0.09) | (0.10) | **(0.10)** ** | (0.09) |
| **Artificial Intelligence -** | -0.02 | 0.01 | -0.07 | -0.13 | **-0.30** | 0.01 |
| **Imperfect Past** | (0.12) | (0.11) | (0.09) | (0.10) | **(0.10)** * | (0.09) |
| **Artificial Intelligence -** | 0.32 | **0.42** | 0.15 | 0.11 | -0.07 | -0.10 |
| **Moral Disagreements** | (0.12) | **(0.12)** ** | (0.09) | (0.10) | (0.10) | (0.10) |
| Biased Worldviews - | 0.16 | 0.12 | -0.06 | -0.13 | 0.03 | -0.02 |
| Difficulty Ethics | (0.12) | (0.11) | (0.09) | (0.10) | (0.10) | (0.09) |
| **Biased Worldviews -** | 0.12 | 0.24 | -0.05 | **-0.38** | **-0.39** | 0.14 |
| **Evil Capacity** | (0.12) | (0.11) | (0.09) | **(0.10)** ** | **(0.09)** ** | (0.09) |
| Biased Worldviews - | 0.03 | -0.00 | -0.04 | -0.22 | -0.16 | 0.14 |
| Imperfect Past | (0.12) | (0.11) | (0.09) | (0.10) | (0.10) | (0.09) |
| **Biased Worldviews -** | **0.38** | **0.41** | 0.18 | 0.01 | 0.07 | 0.03 |
| **Moral Disagreements** | **(0.12)** * | **(0.11)** ** | (0.09) | (0.10) | (0.10) | (0.10) |
| **Difficulty Ethics -** | -0.04 | 0.12 | 0.01 | -0.24 | **-0.42** | 0.16 |
| **Evil Capacity** | (0.12) | (0.11) | (0.09) | (0.10) | **(0.10)** ** | (0.09) |
| Difficulty Ethics - | -0.13 | -0.13 | 0.03 | -0.09 | -0.19 | 0.16 |
| Imperfect Past | (0.12) | (0.11) | (0.09) | (0.10) | (0.10) | (0.09) |
| Difficulty Ethics - | 0.22 | 0.29 | 0.25 | 0.14 | 0.04 | 0.05 |
| Moral Disagreements | (0.12) | (0.12) | (0.09) | (0.10) | (0.10) | (0.10) |
| Evil Capacity - | -0.10 | -0.24 | 0.02 | 0.15 | 0.24 | 0.01 |
| Imperfect Past | (0.12) | (0.11) | (0.09) | (0.10) | (0.09) | (0.09) |
| **Evil Capacity -** | 0.25 | 0.17 | 0.24 | **0.39** | **0.47** | -0.11 |
| **Moral Disagreements** | (0.12) | (0.11) | (0.09) | **(0.10)** ** | **(0.10)** ** | (0.10) |
| **Imperfect Past -** | **0.35** | **0.41** | 0.22 | 0.24 | 0.23 | -0.11 |
| **Moral Disagreements** | **(0.12)** * | **(0.11)** ** | (0.09) | (0.10) | (0.10) | (0.10) |

*Note*: *$p \leq 0.05$; **$p \leq 0.01$

**Experiment 1 Summary**

This study tested five moral humility treatments against a control. Specifically, their impact on moral humility, three political outcomes, epistemic emotions, intellectual humility, and moral relativism was examined. All five treatments —biased worldviews, difficulty ethics, evil capacity, imperfect past, and moral disagreements — significantly increased moral humility. Further, all five interventions were not significantly different from each other. These results together thus suggested that all five treatments can be used as treatments designed to move moral humility (though the strength of the treatments vary across the treatments). Thus, these results served as a kind of a successful manipulation check. Further, they also provided evidence of the amenability of moral humility. However, it is worth noting that the effect sizes were small-to-moderate, suggesting that it would be useful in future work to find ways to amplify the effect of the treatments further.

The interventions' downstream impact on polarization was also assessed across three political outcomes — political sectarianism (capturing partisan animosity), selective exposure (capturing people's tendency towards cross-cutting exposure with the opposing political side), and political compromise (capturing people's willingness to compromise on important political issues). The treatments significantly reduced political sectarianism and increased openness to cross-cutting exposure, but didn't have an impact on political compromise. Specifically, three treatments (biased worldviews, evil capacity, and moral disagreements) significantly reduced political sectarianism, while three (difficulty ethics, evil capacity, and imperfect past) significantly increased interest in exposing oneself to opposing political viewpoints. Notably, across both political outcomes, all five treatments were not significantly different from each other. Mediation analyses largely supported a mechanistic/causal interpretation wherein the treatments increased moral humility, which in turn had a downward impact on political sectarianism and selective exposure in expected ways. However, sensitivity analyses suggested the mediation effect is susceptible to unmeasured confounders.

Two psychological constructs, intellectual humility and moral relativism share the most conceptual similarity with moral humility and have been found to be empirically strongly correlated with moral humility in my previous work. Thus, in this study, the extent to which these treatments, designed to increase moral humility, also affect these two other constructs was examined. The results would give insight into how close these constructs are to moral humility. If they both moved in the same direction and to the same extent as moral humility in response to the interventions, this would suggest that these constructs are very close to each other in the nomological network.

Results indicated that intellectual humility and moral relativism did not move in the same direction or to the same extent as moral humility across these treatments. The results for these two constructs were largely non-significant. Taken together, these provided evidence for moral humility's distinct nature from these two constructs. These constructs perhaps then may be close to each other but maintain their distinctiveness.

Finally, the epistemic emotions experienced by participants were also examined in order to provide insight into and explain any important differences between the five treatments. Broadly, there weren't any pattern of meaningful differences in the emotions (surprise, curiosity, confusion, anxiety, frustration, boredom) evoked by the five treatments. There was some evidence suggesting that the evil capacity treatment evoked more frustration and anxiety compared to other treatments. However, given that the treatments did not perform significantly differently from other treatments when it came to the outcomes, taken in the larger context, evil capacity's negative valence did not change the big picture interpretations and inferences made from the study at this point.

Thus, the overall picture suggested that the five treatments were successful in increasing moral humility and decreasing political sectarianism. It also increased interest in cross-cutting exposure (and its impact on political compromise was inconclusive). These latter two results would although need to be tested in high-powered samples before any strong conclusions can be drawn. Additionally, moral humility appeared to be distinct from intellectual humility and moral relativism. There are some caveats. To reiterate, only the analyses for moral humility, political sectarianism, moral relativism, intellectual humility, and epistemic emotions were highly powered. Other analyses had lower power and thus their results deserve caution in interpretation. Further, while mediation analyses supported a causal pathway, their robustness is weak due to unmeasured confounders. Finally, the effect sizes for moral humility and political outcomes such as political sectarianism and selective exposure were small-to-moderate. This was expected as the interventions were designed as light-touch interventions. However, their practical value might depend on finding ways to enhance the impact of these interventions.

## Experiment 2

Experiment 1 found that the five treatments worked similarly across outcomes but had small-to medium sized effects. Experiment 2 built up on Experiment 1. The main aims of Experiment 2 were twofold: (i) to examine if increasing the dose of moral humility by combining treatments would make the effects on the outcomes stronger, (ii) to replicate the results of Experiment 1, especially given the lower power of some of the analyses in Experiment 1. Experiment 2 had two experimental conditions, a *low-dose* and a *high-dose* condition, and a control condition. The low-dose condition was like Experiment 1 wherein participants randomly received only *one* the five treatments individually. Given that the treatments worked largely similarly in Experiment 1, the five interventions were treated as multiple stimuli instantiating a common construct, allowing to see if the results can be generalized across the five treatments or stimuli. Further, this condition allowed the replication of Experiment 1 as the treatments were presented individually. The *high-dose* condition involved participants receiving a random selection of two of the five treatments in a random order; this served to strengthen the boost in moral humility received by the participants. These two experimental conditions were tested against a control condition. The control was same as Experiment 1 wherein participants read about artificial intelligence. The two experimental conditions were also compared to each other to see if the high-dose condition was significantly stronger the low-dose condition. Finally, mediation of the two treatment conditions on the political outcomes was also examined. All preregistered study details can be found at OSF. Like Experiment 1, Experiment 2 also used a pre-post experimental design to increase precision and power (Clifford et al., 2021) such that outcomes were measured before the treatment as well as after the treatment.

### Dataset and Participants

Experiment 2 was conducted on Prolific. The study was opened to 1500 US participants aged over 18. The sample details are in Table 17. Participants were paid $2.40 for doing the study (~12 minutes). Experiment 2 had two experimental conditions (low and high-dose) and one control condition. A sample of 1500 people (~ 500 in each condition) provided good statistical power (>90%) in a pre-post experimental design to detect a small effect size between each of the experimental and control conditions ($d \sim 0.15$) where the pre and post treatment outcomes are correlated at .7 (for reference, pre- and post-treatment outcomes were correlated at ~ 0.8 - 0.9 in Experiment 1). If and when the pre- and post-treatment outcomes' correlations are higher, the achieved statistical power would be higher. Conversely, if the correlations are lower, the achieved statistical power would be lower.

This sample in Experiment 2 also enabled the detection of a difference of approximately $d \sim 0.15$ between the two experimental groups (high vs. low dose) with strong power (>90%). This means that if the low-dose condition (compared to the control) has an effect size of $d \sim 0.20$, there was >90% power to detect a difference between the low- and high-dose conditions when the high-dose effect is 75% larger (i.e., $d \sim 0.35$). If the high-dose effect is 50% larger than the low-dose effect (i.e., $d \sim 0.30$), the power to detect this difference drops to about 59%. However, for effects exceeding a 75% increase, power approaches 99%. Additionally, if pre- and post-treatment outcomes are more strongly correlated than 0.7, power improves further; for instance, at $r \sim 0.80$, a 50% larger effect in the high-dose condition could be detected with approximately 75% power.

**Materials and Procedure**

***Independent Variable***

Participants were randomly assigned to one of three conditions: the low-dose or high-dose intervention condition, or a control (about artificial intelligence). In the low-dose condition, participants received *one* of the five treatments from Experiment 1 randomly. In the high-dose condition, participants received *two* randomly selected treatments of the five treatments from Experiment 1 in a randomized order.

***Dependent Variables***

A summary of dependent variables included in the experiment is provided in Table 12. Experiment 2 included fewer outcome measures than Experiment 1 to compensate for the increase in the length of the study due to double the number of interventions (in high-dose condition) and more pre-treatment measures (see pre-treatment covariates section below). Specifically, it didn't include measures of intellectual humility, moral relativism, and epistemic emotions as these weren't the main outcomes of interest and were analyzed with good power before. Further, number of political outcomes measured was also reduced to two — with the longest political measure, political compromise removed in Experiment 2.

**Moral Humility.** Moral humility was measured just like Experiment 1, using the 9-item moral humility scale ($\alpha = 0.86$). The impact of the interventions was also explored at the subscale level, i.e., with moral learning/openness ($\alpha = 0.87$), moral fallibility ($\alpha = 0.85$), and moral superiority ($\alpha = 0.77$).

**Political Outcomes.** Two political outcomes were measured. *Political sectarianism* (Finkel et al., 2024) was assessed just like Experiment 1, using the 9-item measure of political sectarianism ($\alpha = 0.96$). The impact of the

interventions was also explored at the subscale level, i.e., othering (α = 0.90), aversion (α = 0.90), and moralization (α = 0.94).

*Selective exposure* was again assessed in the same way as in Experiment 1 wherein participants answered how interested they were in hearing from someone who held the opposing view on an issue that was selected as important by the participant. Accordingly, in Experiment 2, participants were asked to choose only one issue that was important to them which was then was used in the selective exposure task. Like Experiment 1, the importance of the issue and participant's position on the issue were both assessed *before* participants underwent the moral humility intervention.

***Pre-Treatment Covariates***

Experiment 2 also used a pre-post experimental design wherein outcome variables were measured before the treatment as well—which were then treated as covariates in the analysis. Like Experiment 1, the variables included were: moral humility (Moral Humility α = 0.83; Moral Learning/Openness α = 0.83; Moral Fallibility α = 0.78; Moral Superiority α = 0.70) and political sectarianism (Political Sectarianism α = 0.96; Othering α = 0.90; Aversion α = 0.90; Moralization α = 0.94). Additionally, I added a pre-treatment measure of selective exposure (unlike Experiment 1) after observing a small correlation between pre-treatment political sectarianism and post-treatment selective exposure in Experiment 1, which had led to lower power for the selective exposure analyses. I did this because selective exposure was an important outcome, so it was important to enhance the power of its analyses. All three were measured in the same way as their respective post-treatment measures. A summary of pre-treatment measures included in Experiment 2 is provided in Table 12.

***Exploratory Variables***

**Open-Ended Report.** Like Experiment 1, participants reported their general thoughts in an open-ended box at the end of the treatments or control. This was again not a variable of interest and was more of a filler question as part of the cover story of the study and was intended to be analyzed only if time permits.

**Analysis and Results**

The aim of the study was to examine if the low-dose and high-dose moral humility treatments increase moral humility and decrease the polarization outcomes compared to the control condition, and whether the effect of the high-dose treatment is stronger than the low-dose treatment. Additionally, whether the effect of both treatments

on the polarization outcomes is mediated by moral humility was also examined. All preregistered predictions and analyses can be found at OSF.

***Moral Humility***

To test the treatments' effects on moral humility, linear regression was used wherein post-treatment moral humility was regressed on two dummy-coded condition variables and pre-treatment levels of moral humility. Each dummy code compared the low-dose and high-dose treatments to the control (Model 1, Table 29). The prediction was that people assigned to the low- and high-dose treatment conditions will show higher moral humility compared to those in control condition. The correlation between pre- and post- moral humility was $r = 0.85$, $p < .001$, suggesting a high power to detect small effects. Moral humility was significantly higher in the high-dose condition compared to the control condition; however, the low-dose condition was not significantly higher in moral humility compared to control (Model 1, Table 29).

To test if the two experimental groups (low- v high-dose moral humility treatment) significantly differed from each, the reference group in the regression was changed to be the low-dose condition, allowing the comparison of the two experimental groups. Again, pre-treatment level of moral humility was included as a covariate in the analysis. The prediction was that people assigned to the high-dose treatment condition will show higher moral humility compared to those in low-dose condition. The high-dose condition did have a significantly stronger impact on moral humility compared to the low-dose condition by almost 40%. (Model 2, Table 29).

**Table 29**

*Estimates (and standard errors) of the difference in moral humility between each of two treatments (low and high dose) and control (Model 1), and between the two treatments (low vs high dose) (Model 2)*

| | Moral Humility | | | |
|---|---|---|---|---|
| | Model 1 *b* (Reference = Control) | Effect Size Cohen's *d* | Model 2 *b* (Reference = Low Dose) | Effect Size Cohen's *d* |
| High Dose | 0.12** (0.03) | 0.24 | 0.07* (0.03) | 0.14 |
| Low Dose | 0.05 (0.03) | 0.10 | | |
| Control | | | -0.05 (0.03) | -0.10 |
| Moral Humility (Pre-Treatment) | 0.92** (0.02) | - | 0.92** (0.02) | - |
| Observations | 1,540 | | | |

*Note*: *$p \leq 0.05$; **$p \leq 0.01$. Regression coefficients (*b*) are on the original 1-7 scale.

The effect of the low and high dose treatments compared to the control on the three moral humility subscales was also examined. The results for the subscales showed that different aspects of moral humility (moral learning/openness, moral fallibility, moral superiority) contribute to the treatment effect to different extent. The strongest effects were on moral fallibility, followed by moral learning/openness, with negligible effects on moral superiority. Results are in Table 30.

**Table 30**

*Estimates (and standard errors) of the difference in moral humility subscales between each of two treatments (low and high dose) and control*

| | Moral Learning *b* | Moral Fallibility *b* | Moral Superiority *b* |
|---|---|---|---|
| High Dose | 0.14** | 0.19** | -0.04 |
| | (0.04) | (0.05) | (0.05) |
| Low Dose | 0.06 | 0.10 | -0.01 |
| | (0.04) | (0.05) | (0.05) |
| Moral Learning (Pre-Treatment) | 0.88** | | |
| | (0.01) | | |
| Moral Fallibility (Pre-Treatment) | | 0.82** | |
| | | (0.02) | |
| Moral Superiority (Pre-Treatment) | | | 0.82** |
| | | | (0.02) |
| | | 1,497 | |

*Note*: *$p \leq 0.05$; **$p \leq 0.01$. Regression coefficients (*b*) are on the original 1-7 scale.

Moderation analysis was not preregistered but explored. Specifically, I examined if the effect of the low and high-dose treatments on moral humility was moderated by party identity or education to assess if the treatments work similarly across people of different political leanings and educational levels. To test this, a party identity dummy variable (Democrat/Democrat leaning = 0, Republican/Republican leaning = 1) and education dummy variable (less than bachelor's education = 0, bachelor's education and higher = 1) was created and an interaction variable between these and the experimental condition dummy variable was added in the regression. There weren't any significant interactions, suggesting that the treatments worked similarly across political leanings and education levels.

Taken together, the results suggested the high dose moral humility treatment increased moral humility compared to the control, the low-dose treatment did not significantly increase moral humility compared to the control, and the high-dose treatment was significantly stronger than the low-dose treatment. Additionally, both treatments worked similarly for Democrats and Republicans and people of different education levels.

***Political Outcomes***

To test the interventions' effect on polarization, two political outcomes, political sectarianism and selective exposure were examined. A similar test was conducted as moral humility, but on the political outcomes. That is, linear regression was used to test if the two political variables significantly differed between both treatment conditions and control, as well as if the high-dose condition had a bigger impact on the outcomes than the low-dose condition. For political sectarianism analysis, pre-treatment level of political sectarianism was included as a covariate; for selective exposure analysis, pre-treatment levels of selective exposure was included as a covariate. The correlation between pre- and post- political sectarianism was $r = 0.94$, $p < .001$, and the correlation between pre- and post- selective exposure was $r = 0.93$, $p < .001$, suggesting a high power to detect small effects. The prediction was that compared to the control condition, people in the low and high-dose treatments will express lower political sectarianism and more interest in cross-cutting exposure, and that the effects will be stronger in the high-dose condition versus the low-dose condition.

**Political Sectarianism.** Both the low-dose and high-dose interventions did not have a significant impact on political sectarianism (Model 1, Table 31). The high-dose condition was not significantly different from the low-dose condition (Model 2, Table 31).

**Table 31**

*Estimates (and standard errors) of the difference in political sectarianism between each of two treatments (low and high dose) and control (Model 1), and between the two treatments (low vs high dose) (Model 2)*

| | Political Sectarianism | | | |
|---|---|---|---|---|
| | Model 1 *b* (Reference = Control) | Effect Size Cohen's *d* | Model 2 *b* (Reference = Low Dose) | Effect Size Cohen's *d* |
| High Dose | -0.06 (0.04) | -0.09 | -0.04 (0.04) | -0.07 |
| Low Dose | -0.02 (0.04) | -0.03 | | |
| Control | | | 0.02 (0.04) | 0.03 |
| Political Sectarianism (Pre-Treatment) | 0.99** (0.01) | - | 0.99** (0.01) | - |
| Observations | 1,496 | | | |

*Note*: *$p \leq 0.05$; **$p \leq 0.01$. Regression coefficients (*b*) are on the original 1-7 scale.

The impact on the subscales was also not significant. Looking at just magnitude of effects, the results showed that different aspects of political sectarianism (othering, aversion, moralization) contribute to the small

effect of the low-dose and high-dose treatments to different extent (Table 32). The treatments impacted the othering aspect of political sectarianism most, followed by moralization, and then aversion.

Like moral humility before, I conducted moderation analysis (not preregistered, was exploratory) to examine if the effect of the low and high-dose treatments was moderated by party identity or education. There weren't any significant interactions of both with the two treatments, suggesting that the treatments worked (or didn't work) similarly in reducing political sectarianism across political leanings and education levels.

**Table 32**

*Estimates (and standard errors) of the difference in political sectarianism subscales between each of two treatments (low and high dose) and control*

| | Political Sectarianism Othering *b* | Political Sectarianism Aversion *b* | Political Sectarianism Moralization *b* |
|---|---|---|---|
| High Dose | -0.06<br>(0.05) | -0.04<br>(0.04) | -0.06<br>(0.05) |
| Low Dose | -0.03<br>(0.05) | 0.01<br>(0.04) | -0.03<br>(0.05) |
| Political Sectarianism Othering (Pre-Treatment) | 0.95**<br>(0.01) | | |
| Political Sectarianism Aversion (Pre-Treatment) | | 0.98**<br>(0.011) | |
| Political Sectarianism Moralization (Pre-Treatment) | | | 0.96**<br>(0.010) |
| Observations | | 1,496 | |

*Note*: *$p \leq 0.05$; **$p \leq 0.01$. Regression coefficients (*b*) are on the original 1-7 scale.

***Political Sectarianism Mediation*.** The results of the mediation analysis are in Table 33 (left side of the table), wherein I tested the extent to which moral humility explained the effects of the two treatments on political sectarianism. The ACME (Average Causal Mediation Effect) tells us how much of the effect of the interventions on political sectarianism is mediated by moral humility. Thus, it is the indirect effect of the interventions on political sectarianism through moral humility. If ACME is significant (i.e., confidence interval doesn't contain zero), it suggests that moral humility as the mediator plays a significant role in explaining the relationship between the treatments and political sectarianism.

The expectation was that the ACME's will be negative and significant, indicating that the treatments increase moral humility and an increase in moral humility leads to a decrease in political sectarianism. Result

showed that moral humility significantly mediated (ACME) the effect of the high dose interventions on political sectarianism (marginally significant mediation for low dose treatment at $p = 0.06$). The direct and total effects of the interventions on political sectarianism were not significant.

**Table 33**

*Mediation results for the effect of the low and high dose treatments on political sectarianism and selective exposure via moral humility*

| | Political Sectarianism | | Selective Exposure | |
|---|---|---|---|---|
| | Low Dose | High Dose | Low Dose | High Dose |
| ACME | -0.01 | -0.02** | 0.22 | 0.74* |
| ADE | <.01 | -0.03 | 1.25 | 1.90 |
| Total Effect | -0.01 | -0.05 | 1.47 | 2.64 |
| Prop Mediated | 0.79 | 0.35 | 0.15 | 0.28 |

*Note*: †$p \leq 0.1$, *$p \leq 0.05$; **$p \leq 0.01$

I also conducted sensitivity analysis to address the sequential ignorability assumption (Imai et al., 2010; Tingley et al., 2014). The results of this sensitivity analysis showed that across both mediation models, a correlation of $r \sim 0.1$ due to unmeasured confounders could nullify the mediation effect.

Taken together, the results suggested that both the low-dose and high-dose treatments did not significantly reduce political sectarianism, both treatments were not significantly different from each other, and the effects were not moderated by party identity or education levels. Mediation models found some evidence consistent with the hypothesized causal prediction in the high-dose condition, i.e., treatment increases moral humility which consequently decreases political sectarianism. However, sensitivity analysis suggested that the mediation models in general were very sensitive to the existence of confounders.

Why might the mediation effect be significant when the direct and total effects are not? Some authors suggest (Rucker et al., 2011) that the power of the mediation analysis is stronger than the analysis for direct and total effects, making it possible to detect very small effects. Thus, it is possible that there was an effect of the treatments, but given its small size, the tests for direct and total effects (Tables 26 and 27) were nonsignificant; the stronger power of the indirect effects allowed the mediation pathway to reach significance despite the overall weak impact of the treatments. However, give the general patterns of small effects and non-significance, it is hard to tell whether these results are meaningful.

**Selective Exposure.** Both the low-dose and high-dose interventions did not have a significant impact on willingness to engaging with opposing political viewpoint on participant's self-selected most important issue (Model 1, Table 34). The high-dose condition was not significantly different that the low-dose.

**Table 34**

*Estimates (and standard errors) of the difference in selective exposure between each of two treatments (low and high dose) and control*

| | Selective Exposure | | | |
|---|---|---|---|---|
| | Model 1 *b* (Reference = Control) | Effect Size Cohen's *d* | Model 2 *b* (Reference = Low Dose) | Effect Size Cohen's *d* |
| High Dose | 2.74 (1.78) | 0.11 | 1.03 (1.78) | 0.04 |
| Low Dose | 1.71 (1.78) | 0.07 | | |
| Control | | | -1.71 (1.77) | -0.07 |
| Selective Exposure (Pre-Treatment) | 0.92** (0.01) | - | 0.92** (0.01) | - |
| Observations | 1,228 | | | |

*Note*: *$p \leq 0.05$; **$p \leq 0.01$. Regression coefficients (*b*) are on the original -100 - +100 scale.

Like before, I conducted moderation analysis (not preregistered, was exploratory) to examine if the effect of the low- and high-dose treatments was moderated by party identity or education. There weren't any significant interactions of both with the two treatments, suggesting that the treatments worked (or didn't work) similarly in increasing interest in opposing viewpoint across political leanings and education levels.

***Selective Exposure Mediation.*** The results of the mediation analysis are in Table 33 (right side of table), wherein the extent to which moral humility explained the effects of the interventions on selective exposure was examined. The expectation was that the ACME's will be positive and significant, indicating that the treatments increase moral humility and an increase in moral humility leads to increase in interest in opposing political viewpoint. Result showed that moral humility significantly mediated (ACME) the effect of the high-dose, but not the low-dose intervention on selective exposure. The direct and total effects of the treatments on selective exposure were not significant. The results of the sensitivity analysis showed that across both mediation models, a correlation of $r \sim 0.1$ due to unmeasured confounders could nullify the mediation effect.

Taken together, the results suggested that both the low-dose and high-dose treatments did not significantly decrease selective exposure, both the low-dose and high-dose treatments were not significantly different, and the

effects were not moderated by party identity or education levels. Mediation models found some evidence consistent with the hypothesized causal prediction, i.e., treatments increased moral humility which consequently increased cross-cutting exposure in the high-dose condition. However, sensitivity analysis suggests that the mediation models in general were very sensitive to the existence of confounders. Like political sectarianism, it is possible that there was an effect of the treatments but given its pretty small size, the tests for direct and total effects (Tables 28 and 27) were nonsignificant; the stronger power of the indirect effects allowed the mediation pathway to reach significance despite the overall weak impact of the treatments. However, give the general patterns of small effects and non-significance, again it is hard to tell whether these results are meaningful.

***Replication of Experiment 1: Comparing each of the five treatments comprising the low-dose condition to control***

So far, when comparing the *low-dose* condition to the control, all five treatments were analyzed together as one, as the idea was to see if the results found in Experiment 1 generalize over the different moral humility "stimuli" or "topics" (Clifford & Rainey, 2024). This decision was taken in the light of the five treatments behaving similarly in Experiment 1.

However, since participants randomly received one of five moral humility treatments individually in this low-dose condition, it allowed the comparison of each of these five treatments with the control (and with each other). Thus, we could assess if the five treatments were behaving similarly in Experiment 2 as well. Since this analysis was similar to the analysis in Experiment 1, it allowed the replication of Study 1. Because the pre-and post-treatment measures (for all 3 outcomes: moral humility, political sectarianism, and selective exposure) were highly correlated ($r \sim 0.9$), there was good power (>80% to detect $d = 0.15$) even with just approximately 100 participants per the five treatments (biased worldview, difficulty ethics, evil capacity, imperfect past, and moral disagreement) and ~500 participants in control. Thus, each of the five treatments was compared to the control for all three outcomes, moral humility, political sectarianism, and selective exposure. Results are in Table 35.

For moral humility, compared to the control, two treatments worked significantly to increase moral humility and had comparable effect sizes like Experiment 1. These were the biased worldviews and moral disagreement treatments. Thus, these two worked consistently across Experiment 1 and 2. The other three treatments had either negligible and non-significant effect (difficulty ethics, past treatments), or opposite of the expected effect (evil capacity, although the effect was not significant). Thus, these latter three didn't replicate across the two experiments.

**Table 35**

*Estimates (and standard errors) of the difference in moral humility, political sectarianism, and selective exposure between each of five individual treatments comprising the low dose condition and control*

| | Moral Humility *b* | Moral Humility Cohen's *d* | Political Sectarianism *b* | Political Sectarianism Cohen's *d* | Selective Exposure *b* | Selective Exposure Cohen's *d* |
|---|---|---|---|---|---|---|
| Biased Worldviews | 0.18** (0.05) | 0.36 | -0.08 (0.06) | -0.15 | 2.42 (2.96) | 0.10 |
| Difficulty Ethics | 0.02 (0.05) | 0.04 | -0.001 (0.06) | -0.003 | 1.98 (3.01) | 0.08 |
| Evil Capacity | -0.06 (0.05) | -0.12 | 0.02 (0.06) | 0.03 | 3.94 (3.17) | 0.16 |
| Imperfect Past | 0.001 (0.05) | 0.001 | 0.15** (0.06) | 0.26 | -6.14* (3.15) | -0.24 |
| Moral Disagreements | 0.11* (0.05) | 0.23 | -0.15* (0.06) | -0.27 | 5.65* (2.97) | 0.22 |
| Moral Humility (Pre-Treatment) | 0.94** (0.02) | - | | | | |
| Political Sectarianism (Pre-Treatment) | | | 0.99** (0.01) | - | | |
| Selective Exposure (Pre-Treatment) | | | | | 0.92** (0.01) | - |
| Observations | 1,025 | | 997 | | 823 | |

*Note*: *$p \leq 0.05$; **$p \leq 0.01$

For political sectarianism, compared to the control, again the same two treatments replicated in terms of significance or comparable effect sizes in decreasing political sectarianism. They were the moral disagreement and biased worldview treatments. The other three treatments had either negligible and non-significant effect (difficulty ethics) or opposite of the expected effect (imperfect past and evil capacity, former was stronger and significant). Thus, these latter three didn't replicate across the two experiments. So far across the outcomes, looking at direction of coefficients, two treatments, evil capacity and imperfect past seemed to be working in opposite of expected direction. Evil capacity seemed to work in opposite of intended direction for moral humility and political sectarianism, and imperfect past in opposite direction for political sectarianism.

For selective exposure, compared to the control, one treatment worked in opposite of expected direction to significantly decrease interest in exposure to opposing political viewpoint. This was the imperfect past treatment which also worked in opposite of the expected direction for political sectarianism. The other four broadly replicated in terms of significance or comparable effect sizes in increasing interest in exposure to opposing political viewpoint.

However, some were stronger and significant than others (i.e., moral disagreement), although the four were not significantly different from each other.

In sum, across Experiment 1 and 2 taken together, the overall picture from the three outcomes suggested that some treatments worked more consistently, strongly, and as hypothesized (e.g., biased worldviews, moral disagreements), some worked inconsistently and at times produced effects in the opposite direction than hypothesized (e.g., evil capacity and imperfect past), and some worked in the direction as expected but had very weak effects (e.g., difficulty ethics).

***Exploratory Analyses***

**Exploratory Analysis 1: Zooming in on the treatments in the high-dose condition.** Some of the five treatments in the low-dose condition, i.e., imperfect past and evil capacity, worked in the opposite of the expected direction on the outcomes when comparing each of them individually to the control. Given this, I next did exploratory analyses (not preregistered) examining if the presence of these two treatments in the high-dose condition, wherein they were randomly presented with one other treatment had counteractive effects. That is, whether the presence of these two (imperfect past and evil capacity) treatments led to a diminished effect or an effect in the opposite direction to the expected effect on the three outcomes. I compared each of the ten possible treatment combinations to the control condition (five treatments yield ten possible combinations). Since these were exploratory analyses with reduced power, I focused on the sign and magnitude of the effects and not the significance.

Across the ten possible combinations of the five treatments, I did find evidence suggestive of the counteractive effects of these two treatments (Table 36). For example, the combination of these two treatments produced the smallest positive effect on moral humility, or whenever there was increased political sectarianism (positive sign) or decreased willingness towards exposure to opposing viewpoints (negative sign), the treatment combination included the presence of imperfect past treatment.

Thus, when considered alongside the previous analysis of the five treatments individually in the low-dose condition wherein evil capacity and imperfect past were behaving oddly, the findings suggest that the imperfect past and evil capacity treatments tend to be backfiring at times and possibly be having counteractive effects. Thus, it is possible that their presence maybe obscuring or interfering with the effects of other treatments when analyzed together in both the low-dose and high-dose conditions.

**Table 36**

*Estimates (and standard errors) of the difference in moral humility, political sectarianism, and selective exposure between each of the ten possible combinations of the five treatments in the high dose condition compared to control*

| Contrast (the treatment combination vs control) | Moral Humility *b* | Political Sectarianism *b* | Selective Exposure *b* |
|---|---|---|---|
| difficulty ethics & imperfect past | 0.13 (0.07) | 0.07 (0.08) | 8.44 (3.87) |
| difficulty ethics & biased worldview | 0.07 (0.07) | -0.13 (0.09) | 9.73 (3.83) |
| difficulty ethics & evil capacity | 0.16 (0.07) | -0.14 (0.09) | 0.44 (3.91) |
| difficulty ethics & moral disagreements | 0.10 (0.07) | -0.06 (0.09) | 4.73 (4.26) |
| imperfect past & biased worldview | 0.20 (0.07) | 0.08 (0.09) | -3.69 (3.79) |
| imperfect past & evil capacity | 0.03 (0.07) | -0.11 (0.09) | -5.02 (4.21) |
| imperfect past & moral disagreements | 0.06 (0.07) | 0.03 (0.09) | 4.65 (3.87) |
| biased worldview & evil capacity | 0.15 (0.07) | -0.06 (0.09) | 3.96 (4.00) |
| biased worldview & moral disagreements | 0.14 (0.07) | -0.13 (0.09) | 3.14 (3.87) |
| evil capacity & moral disagreements | 0.14 (0.07) | -0.10 (0.09) | 0.39 (3.83) |

**Exploratory Analysis 2: Reanalyzing the main models without the two counteractive treatments.**

Next, I reanalyzed the main models for Experiment 2, excluding these two treatments from both the low-dose and high-dose conditions. Specifically, I examined whether the low-dose and high-dose conditions increase moral humility, decrease political sectarianism, and increase interest in cross-cutting exposure compared to the control condition, and whether high-dose conditions have stronger effects than low-dose conditions on the three outcomes. Results are in Table 37.

Both low- and high-dose conditions significantly increased moral humility compared to the control. The high-dose condition significantly decreased political sectarianism and significantly increased selective exposure compared to the control. The low-dose condition marginally decreased political sectarianism ($p = 0.05$) and increased selective exposure ($p = 0.07$) compared to the control. Comparing the low- and high-dose condition to each other, the three outcomes were not significantly different across the two conditions—meaning the high-dose treatment didn't have a meaningfully stronger effect on outcomes than the low-dose treatment.

**Table 37**

*Estimates (and standard errors) of the difference in moral humility, political sectarianism, and selective exposure between each of two treatments (low and high dose) and control, excluding two treatments (evil capacity and imperfect past)*

| | Moral Humility | Effect Size Cohen's *d* | Political Sectarianism | Effect Size Cohen's *d* | Selective Exposure | Effect Size Cohen's *d* |
|---|---|---|---|---|---|---|
| High Dose | 0.10$^{*}$ (0.04) | 0.21 | -0.11$^{*}$ (0.05) | -0.20 | 6.01$^{**}$ (2.48) | 0.25 |
| Low Dose | 0.10$^{**}$ (0.03) | 0.22 | -0.08$^{\dagger}$ (0.04) | -0.15 | 3.36$^{\dagger}$ (1.88) | 0.15 |
| Moral Humility (Pre-Treatment) | 0.95$^{**}$ (0.02) | - | | | | |
| Political Sectarianism (Pre-Treatment) | | | 0.99$^{**}$ (0.01) | - | | |
| Selective Exposure (Pre-Treatment) | | | | | 0.93$^{**}$ (0.01) | - |
| Observations | 938 | | 941 | | 791 | |

*Note*: $^{\dagger}p\leq0.1$, $^{*}p\leq0.05$; $^{**}p\leq0.01$

**Experiment 2 Summary**

This study tested a low-dose and high-dose moral humility intervention against a control and each other. Specifically, their impact on moral humility, political sectarianism, and selective exposure was examined. In the low-dose intervention, participants received one of five treatments randomly, and in the high-dose intervention they received two of five treatments randomly. Only the high-dose intervention significantly increased moral humility. Neither intervention significantly reduced political sectarianism or increased willingness to engage with opposing political viewpoints. Further, there were no significant difference between the two interventions on these outcomes (except for moral humility where the high-dose had stronger impact than the low-dose). Mediation analyses indicated that the interventions influenced political sectarianism and selective exposure indirectly through moral humility in the high-dose condition. Taken together, given that these indirect effects were small, the direct and total effects were nonsignificant, and in general most models were non-significant, the overall impact of the interventions was uncertain.

A closer examination of the five individual treatments that comprised the low-dose and high-dose conditions – difficulty ethics, biased worldviews, imperfect past, evil capacity, and moral disagreements – revealed

that there was a notable variation in their effect. It appeared that the five treatments were not working similarly, unlike Experiment 1. The most consistent and overall strongest and significant treatments across outcomes and experiments were the biased worldviews and moral disagreements treatments. The imperfect past and evil capacity treatments were sometimes working in the opposite direction than expected in Experiment 2 and thus were inconsistent across the two experiments. The difficulty ethics treatment seemed to have very negligible effects at times, although overall behaved in the predicted direction across outcomes and experiments.

Taken together, this means that combining the five treatments or treating them as largely similar or interchangeable isn't the most effective approach at this stage as their individual impacts vary significantly. Doing this for the main analysis in this study might have obscured effects when they existed and/or distorted the effects due to the inter-stimuli/treatment variation. To address this concern to some extent, I did exploratory analysis, re-estimating the main models with the two unreliable treatments excluded. This revealed that the low-dose and high-dose conditions did significantly (sometimes marginal) increase moral humility, decrease political sectarianism, and increase cross cutting exposure. However, the high-dose condition was still not significantly stronger than the low-dose condition; they both had very similar effect sizes. Thus, combining treatments didn't have a substantial compounding effect in increasing the strength of the interventions that I had expected. In a similar vein, overall, the effect sizes for moral humility and political outcomes were again in the small-to-moderate range like Experiment 1.

In summary, these findings suggested that while the treatments had some impact, their effectiveness was modest and varied depending on the specific treatments used. Of the five treatments used in both experiments, why might have some treatments worked more consistently (like biased worldviews or moral disagreements) than others which sometimes even backfired (like evil capacity and imperfect past)? One possibility is that some treatments like the latter two are more sensitive to context. Notably, the political context changed between the two experiments—the change in government between the two experiments created a more politically charged atmosphere. At a time when political identities were highly charged and the political climate intensely moralized, interventions that highlighted the capacity of people to do evil like erstwhile Nazis, or compared the moral failing of present generations to those of their ancestors might have triggered or threatened participants. This could have occurred if these elements in the treatments were perceived as an attack on their ingroup's identity or worldview or as validating moral disdain towards outgroups or vindicating their own ingroup's worldview. It is also possible that since these two treatments had the possibility of evoking negative emotions of guilt and shame given that they both talked about

wrongdoings of past groups and ancestors, they are more susceptible to backfiring if these emotions trigger moral identity threat. Future work would thus need to investigate these possibilities and refine these treatments to minimize unintended backfire effects.

## Chapter IV: General Discussion, Future Directions, and Conclusion

In this project, across ten studies, I investigated the psychological construct of moral humility, filling existing gaps in its empirical study. I conceptualized and developed a psychometrically validated measure, examined its nomological associations, and analyzed its antidotal role in a moralized and conflictual context, i.e., political polarization, using correlational and experimental studies. Notably, to this latter end, I developed and validated interventions of moral humility and assessed their causal impact on polarization outcomes. This also enabled the examination of a more fundamental question—whether moral humility an individual characteristic can be changed or is relatively stable.

I constructed a thirty-item moral humility scale comprised of three factors — moral fallibility, moral openness/learning, and moral superiority. An examination into its nomological associations, especially to constructs such as personality, religiosity, ideology, political extremity, moral grandstanding, moral relativism, and intellectual humility, showed that moral humility was associated with these constructs in meaningful ways. A person high in moral humility showed less extremism and absolutism (e.g., less political extremity, religious exclusiveness, moral grandstanding, moral absolutism, religiosity), more openness and flexibility (e.g., openness to experience, intellectual humility, need for cognition), more humility (e.g., modesty, intellectual humility), and a more positive orientation towards others (e.g., more agreeableness, lower psychopathy, lower narcissism, etc.).

Having constructed a scale of moral humility and assessed that it was working sensibly with other constructs, I then assessed its ability to attenuate the dark features of morality in a moralized and high-conflict context. In the context of political polarization in the US, moral humility predicted lower antagonism and derogation towards outgroup, lower rigidity in one's own views, lower rejection of compromise and contact, and lower adoption of morally questionable means. Consistently, across over fifteen outcomes and multiple studies in diverse samples, moral humility was associated with lower levels of polarization. These relationships provided strong evidence of moral humility being a counteractive force in high-conflict, intergroup contexts that bring out the dark side of our morality. Further, the studies found that these effects of moral humility emerged over and above effects of other important, conceptually and empirically related constructs such as intellectual humility and moral relativism. However, these results were correlational in nature and could not definitely establish the causal role of moral humility.

Five moral humility interventions were then designed and tested in two experiments. These were aimed to find ways to increase people's moral humility. They thus served to provide insight into the amenability of moral humility (i.e., if it is a fixed or changeable aspect of people), which would help shed light on whether it is a construct that can be fruitfully targeted through interventions. They also helped examine the causal role played by moral humility in a moralized, conflictual contexts like political polarization. Across the two experiments, two interventions, biased worldviews and moral disagreements, emerged as the most consistent treatments that increased moral humility, and subsequently decreased political antagonism towards outgroup and increased openness towards opposing sides' viewpoints. A third treatment, difficulty ethics also worked in the predicted direction across outcomes but sometimes had negligible effects. Finally, two interventions, evil capacity and imperfect past, produced inconsistent effects across experiments. These variations suggest that the approach taken in the second experiment, where interventions were treated as interchangeable or combined, should be avoided until their individual consistency is better established. Nonetheless, the three more reliable interventions provided promising evidence that moral humility can be increased (although with modest effect sizes), which can then have ameliorative effects in morally inflamed contexts such as polarization. These findings also highlight the kind of ideas (e.g., biased worldviews, moral disagreements, difficulty ethics) that show promise for boosting moral humility (either individually and/or in combination) and counteracting the rigid and antagonistic tendencies that arise in moralized conflicts. Further, by showing the amenability of moral humility, they supported the value in targeted interventions to increase moral humility. The modest effectiveness of these interventions however does raise the need for further refinement of these intervention strategies to enhance their robustness and impact.

## Theoretical and Practical Implications

### *Moral humility as a distinct and meaningful psychological construct*

Morality is a fundamental axis of human experience, shaping how our individual selves and social lives are understood and navigated. Notably, it has a dark side that fuels hate, intolerance, rigidity, cruelty, violence, and conflict. Against this backdrop of morality's centrality to our lives and its dark nature, the main thesis of this project was to propose and show that a domain-specific form of humility—moral humility, or humility in the domain of morality—is a conceptually and empirically meaningful psychological construct for understanding our morally infused lives. In particular, this project showed that moral humility can serve as an antidote to the dark features of our morality, for instance, as they manifest in a polarized, moralized, and conflictual intergroup (political) context.

Moral humility as a domain-specific form of humility has not received much conceptual or empirical examination. This work puts moral humility on the map, as a distinct, theoretical, and meaningful psychological construct, worth the investigation, especially when it comes to understanding morally relevant contexts and psychological tendencies.

Relatedly, using correlational or/and experimental evidence, this project demonstrated moral humility's uniqueness from important constructs such as general humility (in HEXACO), intellectual humility, moral conviction, and moral relativism. This is important, as it suggests that a distinct form of humility—moral humility—plays a unique role in navigating morally charged situations, offering a psychological mechanism that is separate from general or intellectual humility, or moral conviction or relativism. This distinction opens the door for further research into how moral humility influences key outcomes, especially those that are morally relevant. In particular, when it comes to countering the dark aspects of our morality, there are few works so far in the moral psychological literature, directly or indirectly providing solutions that put morality at its core (for exceptions see Kraaijeveld & Jamrozik, 2022; Kodapanakkal et al., 2022; Wright & Pölzler, 2022; Bastian et al., 2015). This work fills this gap and provides an additional promising solution to the repertoire of possible antidotes.

Notably, amongst the solutions suggested in existing literature, two are de-moralization or moral relativism (Kraaijeveld & Jamrozik, 2022; Kodapanakkal et al., 2022; Wright & Pölzler, 2022; Bastian et al., 2015). However, moral conviction or moralization has also been associated with positive outcomes such as voting or collective action (Skitka & Bauman, 2008; Van Zomeren at al., 2012). It is thus unclear if reducing moral conviction would compromise its positive effects, and hence whether doing so would be advisable. Moral relativism has been considered philosophically and morally objectionable by many scholars due to its implications of "anything goes" wherein harmful practices such as sexism can be justified as being right according to the norms of a sexist society (Gowans, 2021). Given these competing considerations for both these proposed antidotes, moral humility offers the potential of an underexplored psychological antidote that does not rely on demoralization or inducing moral relativism to counteract our dark moral tendencies. It would instead encourage approaching moral divisions with an acknowledgement of one's own moral limitations, recognition of others' moral strengths, and an openness to moral learning. Of course, at this stage, there is a lack of definitive comprehensive data to make strong claims about moral humility's unequivocal positive effects without also investigating its secondary effects. Thus, any strong claims await future tests. Nevertheless, the evidence in this project far points towards a promising solution to the dark aspects of our morality.

### *Moral basis of political conflicts*

When it comes to political polarization and political conflicts at large, this work offers additional insight and support into its moral basis. Politics is not merely about practical concerns but also has moral and value considerations at its core (Paul et al., 2010; Colombo, 2019). These considerations are often given less attention when designing interventions as compared to political identity (e.g., Voelkel et al., 2024), despite evidence of its close connections to people's political attachments (Lupton et al., 2020). Focusing on political identity is understandable; however, because political identities are largely stable overtime (more stable than personality traits; Brandt & Morgan, 2022; Vaisey & Kiley, 2021), identifying routes to depolarization that do not invoke political identities is fruitful. By showing moral humility's relevance to political polarization, the work in this project a) reinforces the moral nature of politics and political conflicts, and b) offers a morality-based approach to addressing political conflicts. Relatedly, it shows using experimental studies that moral humility is amenable to change, offering a way to addressing conflicts that is not focused on shifting stable psychological attributes like political identity. Moreover, in the correlational studies on political polarization, moral humility's effect size was on average greater than that of intellectual humility and political extremity, two psychological constructs that have been investigated in the context of polarization, and at least one that is considered central (i.e., political extremity). This suggests that moral humility is likely of substantive importance. Taken together, this research on moral humility and political polarization highlights moral humility as a promising and underexplored avenue for addressing political polarization—one that acknowledges the moral underpinnings of political divisions.

## Strengths, Limitations, and Future Directions

The work in this project had the following strengths. First, it developed the first psychometrically validated measure of moral humility, allowing future studies to measure this construct and build upon it. Relatedly, in addition to the first validated moral humility scale, it also offers a shorter scale of moral humility for use in large national and expensive surveys (used in some studies here). Second, by providing the first comprehensive examination into moral humility's nomological network, it provides a solid groundwork on the nature and correlates of the construct which future work can build on. Third, it is amongst the first to show the relevance of moral humility in a high-conflict, polarizing, and moralized domain, i.e., politics. This highlights the importance of understanding this construct in the context of political outcomes and contexts. Notably, the relationship between moral humility with polarization emerged across multiple outcomes, ranging from hostility towards outgroup, perceptions of outgroup morality and

threat, willingness to politically compromise, support for anti-democratic methods, and interest in cross-cutting exposure. Fourth, it is amongst the first to develop and validate interventions to increase moral humility, showing that moral humility is responsive to targeted interventions and the kinds of ideas that might help in boosting moral humility. This provides rationale and motivation for further developing interventions that increase moral humility even more strongly and in lasting ways in future work by amping up the moral humility treatments (such as by repeating intervention over time or embedding them in public platforms or conversations). Fifth, this work was conducted across different types of samples (YouGov nationally representative samples, student samples, nationally diverse online Prolific samples), with relationships replicating across the diverse samples. The use of both diverse samples and outcome lends confidence to the reliability and generalizability of the findings. Sixth, this work provided important evidence (through incremental and discriminant validity tests) to distinguish moral humility from related constructs like intellectual humility and moral relativism, suggesting an alternative psychological mechanism through which important and relevant outcomes can be targeted. Seventh, moral humility's effect size in the correlational studies being on an average comparable or greater than intellectual humility and political extremity—two psychological constructs that have been investigated and considered important in the context of polarization—showed that moral humility is actually of substantive importance. Finally, this work provided both correlational and causal evidence towards supporting moral humility's role in mitigating the dark features of our morality, particularly in the context of political divisions. Overall, this work is amongst the first detailed empirical investigations into moral humility, its nature, associations, and its impact.

Despite these strengths, there were limitations. First, the relevance of moral humility in countering the dark features of morality was examined in just one morally inflamed conflict i.e., political polarization. Thus, we have limited insight into its relevance in other moralized contexts. Future work should investigate the role of moral humility in other contexts characterized by these dark moral features such as religious intolerance or historical conflicts. Further, to fully understand the nature and consequences of moral humility, future studies also need to examine how moral humility is linked with the not-so-dark aspects of morality. For example, how is moral humility associated with taking action against moral wrongs or fighting against injustice? The current studies showed that moral humility can counter the dark aspects of our morality such as animosity and rigidity, but can moral humility also promote the constructive aspects of our morality? Or does moral humility dampen moral action, such as social activism aimed at moral progress and social change? The answers to these questions, and consequently a fuller

understanding of moral humility, await further investigation. Notably, some previous work has found that observing moral humility in others promotes prosocial and ethical action (Owens et al., 2019), suggesting that moral humility may promote the good features of morality as well.

Second, the measures used across studies were largely self-report. While self-report measures have notable strengths (Corneille & Gawronski, 2024), incorporating behavioral assessments can provide a richer understanding of a psychological phenomenon. Future research on moral humility would benefit from such measures. For instance, researchers could examine how moral humility influences resource allocation in economic games (e.g., ultimatum or dictator games) for those who participants perceive to be on the opposing side in a moral situation and those on the same. Another approach could involve observing people's punishment or forgiveness behaviors in morally ambiguous situations where the incentive for virtue signaling or moral grandstanding is high (e.g., Jordan & Kteily, 2022). Alternatively, willingness to sign up for or attend events like talks or discussions that challenges people's own moral ideas or worldviews could offer valuable behavioral insights into moral humility. Apart from using behavioral measures to enhance the understanding of moral humility, self-report measures that capture state-variations in moral humility across different tasks or moral contexts can enhance the understanding of moral humility by providing insight into its situational dynamics and help triangulate on it using an alternate method (Fleeson & Jayawickreme, 2015; Van Tongeren et al., 2023). The measure developed and used in this project assessed moral humility as a trait or global characteristic of the person. By using state measures of moral humility in future studies, researchers could better capture the variability in moral humility, i.e.., how moral humility fluctuates within and across individual when they are placed in various moral situations.

Third, across the correlational and experimental studies, moral humility was observed to have small-to-medium effects on the outcomes assessed (Funder & Ozer, 2019). This means that while moral humility appears to be a meaningful psychological construct with real-world implications, its impact may be modest in magnitude. Given this, future research can explore ways to enhance the effectiveness of moral humility interventions. Interventions could be made more immersive or sustained over longer periods to see if stronger and more lasting effects can be achieved. For example, repeated exposure to the moral humility-promoting narratives identified in the experiments, or its integration into social conversations with peers or in messages by authority figures, such as leaders or media influencers, could help reinforce its impact. Additionally, examining whether moral humility interacts with other psychological traits (e.g., personality, social identity, character virtues, cognitive and emotional

styles) or contextual factors (e.g., elite signaling, social norms) could also help identify conditions under which its influence is amplified.

Fourth, while the experimental studies did find that moral humility can be increased through certain interventions and reduce political divisions, not all interventions that were tested as part of the studies worked reliably. Two interventions, one emphasizing the capacity for evil in ordinary humans, and one highlighting the moral imperfections and blindspots of our ancestors, produced inconsistent effects across the studies. In the second study, they worked counter to expectations on moral humility or political outcomes. Why might this have happened? One possibility is that these interventions are particularly sensitive to contextual factors. The first and second experimental studies which tested these interventions were conducted at different times, with a major political shift occurring between them—Donald Trump's inauguration and subsequent executive orders substantially changed the political landscape. It is possible that the heightened political tensions at the time of the second study made participants react more defensively to these interventions. This might have especially happened with these two interventions because perhaps they have some elements or ideas that have more susceptibility to backfire effects if perceived as an attack on their ideological group or moral worldview. Future work would thus need to repeatedly test these as well as other interventions across different contexts and time points to establish their robustness. Further, before repeated testing across contexts, these interventions might need to be refined more through more detailed discussions and/or testing with diverse population subgroups, so that any threatening aspects may be identified, removed and/or reframed.

Fifth, while this work showed that moral humility can be increased through targeted interventions, the durability of these effects remains an open question given that the experimental studies in this project did not follow-up participants after the experiments. Do the observed increases in moral humility persist over time, or do they fade as the moral humility narratives fade from salience? Longitudinal studies that track moral humility across extended periods would help determine the stability of these changes in response to interventions.

Sixth, this project focused primarily on moral humility in the political-cultural context of the US. Future research should explore moral humility in different cultural contexts within and across different societies. Given that the conceptions of morality and/or humility may vary across cultures, moral humility may manifest differently in different societies and cultures with different dominant moral frameworks or understanding of humility. For example, consider two religious cultures, Christianity and Hinduism. In Christianity, moral fallibility of humans is

central to its religious and ethical teachings—moral humility in someone influenced by the Christian framework thus may dominantly take the form of recognizing the inherent moral limitations of oneself and others. Alternatively, Hinduism embraces moral particularism as an important part of its moral framework — thus moral humility in someone influenced by the Hindu framework may dominantly take the form of recognizing that moral priorities and strengths may be different across individuals embedded in different contexts. These differences in the source and nature of moral humility might mean that interventions and measurements would need to be adapted to account for these cross-cultural variations. It would also mean refining the theory to account for variations in the manifestations of moral humility.

Finally, insight into the ontogeny of moral humility is lacking and would be useful in understanding the antecedents of moral humility. Future work can investigate how virtues like moral humility develop. Are they primarily shaped during childhood, alongside the development of other core aspects of our morality (e.g., empathy, fairness, or prosocial behavior), or do they emerge more prominently in adulthood, when we become better at reflection and recognizing complexity? Are they influenced by certain religious, cultural, or political values, or by specific life experiences such as exposure to diverse people and perspectives? Insights into these questions await future research and would enrich our understanding of how the virtue of moral humility can be cultivated.

***Digging deeper into unexpected findings***

Throughout this project, there were some unexpected findings that highlighted gaps in the theoretical and practical understanding of moral humility and would benefit from future enquiry. First, while probing moral humility's associations with other psychological constructs and political outcomes, I discovered an unexpected small positive association of moral fallibility with psychopathy and support for anti-democratic means, and small negative associations with sincerity and fairness (facets of honesty-humility). All these constructs broadly captured an inclination towards morally questionable things. Taken together, these associations indicate that acknowledging one's moral limitations or fallibility is not unequivocally a sign that one is motivated towards moral improvement, prosociality, or generally towards thinking or acting morally. Instead, these associations suggest that recognizing one's moral imperfections could, for some individuals, be compatible with an immoral, amoral, or morally complex orientation. For example, imagine an ambitious person (perhaps a politician, businessman, or scientist) who wants to make a tangible impact on society. This person might think that sometimes doing things that are immoral (e.g., exploiting workers, flouting ethical guidelines) are a necessary cost that one must pay to achieve great things. Such a

person might be willing to acknowledge that they indeed do morally questionable things which makes them morally imperfect, but they are not inclined to be any different as they are motivated by other ends. As another example, imagine a person who recognizes that eating meat is morally problematic and doing so makes one morally compromised as a human being—yet they are not inclined to act differently because being a moral person or being moral in the domain of food choice is not very important to their identity or self-concept. Or consider another person who views moral fallibility as inevitable and, based on that perspective, makes little effort to change or grow morally. They may acknowledge moral shortcomings as an unchangeable aspect of human nature and adopt a passive or resigned stance toward their own moral failings, or perhaps even use it as a license to be morally lax. All these examples highlight how an acknowledgment of moral fallibility or limitations is not necessarily an indication that it will promote a more moral, ethical, or prosocial approach in someone.

This raises important questions about the boundary conditions of the implications of moral fallibility, and perhaps moral humility generally. Does moral fallibility operate differently in the presence of other psychological factors? For example, does the acknowledgement of one's moral fallibility produce different behaviors and inclinations when it occurs in the presence of a stronger vs weaker moral identity (i.e., how important being moral is to one's self concept) or higher versus lower levels of moral openness/learning (e.g., another facet of moral humility). Does high moral fallibility look different in a person high in psychopathy versus someone high in empathy? These suggest the need for future work to investigate moral fallibility (and by extension, moral humility) in an integrative fashion to get a nuanced understanding of the nature and implications of moral humility, rather than assuming moral fallibility to be an unequivocally and uniformly a positive force. Broadly, these insights apply not to just moral humility but to the study of humility in general (including other types of humility in literature). An acceptance of one's fallibility is a core part of the conceptualization of humility and is often *assumed* to produce good interpersonal and intergroup outcomes. The theorizing needs to account for cases where the acceptance of fallibility occurs in people who don't care about being wrong or flawed or consider fallibility inevitable, as these kinds of cases might mean differential predictions of fallibility's consequences based on maybe the presence of other psychological characteristics. Moral fallibility is then perhaps a necessary but not sufficient condition for constructive interpersonal and intergroup outcomes.

The second set of unexpected findings emerged when probing the relationship between moral humility and polarization outcomes, specifically its lack of association with myside sharing in a correlational study. I had

anticipated that higher moral humility would be associated with lower sharing of moralized and ideologically congruent news items on social media. Although the results tended in that direction (marginally significant, weak effects), moral humility's role remains unclear. Given that myside sharing was one of the few outcomes across the many studies and outcomes in this project that was more behavioral in nature, these results highlight the important need to understand moral humility's impact on behavior. The studies in this project did provide strong evidence of moral humility's relationship with more cognitive (i.e., learning) and emotional (e.g., antipathy) outcomes, but information on its relationship with behavioral outcomes is lacking. The relationship between self-reports and behavior is famously weak (Dang et al., 2020). One reason for this is that behavior is multi-determined whereby other determinants of the behavior can weaken the relationship between certain thoughts/feelings/attitudes and behavior. In the case of myside sharing, the public or social context of sharing news items on social media might bring other strong contextual considerations in mind (e.g., social condemnation or reward) that trump or weaken the effects of one's moral humility. This possibility highlights that the effects of moral humility on behavior might be context dependent. Thus, it is essential in future work to systematically examine what kinds of contexts can hinder or facilitate the impacts of moral humility. For example, is moral humility less influential in environments where the norms favor moral outrage (e.g., social media), or more influential in contexts where norms favor seeing the good in others (e.g., community or interpersonal dialogues).

In a similar vein, another possible reason for the weak links with behavior is that self-report scales (including mine) are typically designed to be context-insensitive, i.e., they capture the participant's understanding of themselves across contexts (e.g., how morally humble they are across moral situations). Said differently, they are designed to capture traits rather than states. Thus, these trait-oriented scales are sometime poor at predicting behavior in a particular context. One way to address this is to adapt the scales to be more context-sensitive so that they can capture moral-humility states (e.g., how morally humble people think they are when it comes to interacting with people on social media). Thus, this raises the need for self-report scales that can capture states as well, which might then help with predicting morally relevant behavior in particular contexts with greater fidelity. However, it is possible that the lack of relationship with myside sharing was not because of the scale's context-insensitivity but because the current scale captures people's cognitive orientations more than it captures their behavioral orientations. Although during scale development, I aimed to have items that captured behavior (e.g., moral choices, decisions) as well as cognitive orientations (e.g., moral views), in the process of refining the scale, the scale ended up being more

dominated by items capturing the latter. Therefore, even a trait-focused moral humility scale that includes more balanced items measuring both moral behaviors and moral views could be more effective at predicting behavioral outcomes. This would involve refining the scale in the future to more comprehensively capture the full scope of moral humility content.

Finally, the third set of unexpected findings was the inconsistency of certain moral humility treatments or interventions across the two experiments. Some moral humility treatments, contrary to my expectations, were weaker in their effects when tested in a different experiment later or even backfired. As mentioned previously, one possibility is that the increasingly charged political landscape during Experiment 2 caused participants to react more defensively or rigidly to treatments that could be perceived as either an attack on or a validation of a particular political group. Regardless of the exact reason, these findings highlight a crucial gap in the present research—i.e., in order improve the reliability and effectiveness of moral humility interventions, it is important to gain a deeper understanding of the psychological mechanisms driving moral humility and its change. While the intervention studies in this project were primarily outcome-oriented, i.e., focusing on ways to change moral humility, the next phase of research can complement the understanding of moral humility by taking a more process-oriented approach that involves understanding the psychological processes through which these changes occur. For example, understanding how cognitive processes like cognitive reappraisal of one's moral limitations or perspective taking of other's moral strengths, or emotional responses like empathy, guilt, or shame, mediate the effects of interventions could provide insight into why some approaches work better or worse than others. Additionally, exploring how identity dynamics, such as moral or ideological identity/worldview threat or salience influences responses to moral humility interventions may help reveal why certain individuals or groups react negatively or fail to benefit from interventions. By identifying these underlying psychological processes, future research can identify and design interventions that are reliable and effective across different conditions (e.g., individuals, groups, and contexts).

## Conclusion

While the "better angels of our nature" may guide us toward moral progress, there is a dark side to our morality as well that has long fueled violence, conflicts, hate and, intolerance. This project proposed moral humility as a possible antidote to these destructive tendencies of our morality. It offered the first comprehensive empirical examination into the psychological nature and significance of moral humility. Through both correlational and experimental studies, the studies showed moral humility's role in mitigating the rigidity and hostility that often characterize moralized political conflicts. This research also laid the groundwork for future inquiries into moral humility—amongst others, its relationship with moral action and moral improvement, its developmental pathways, strategies for cultivating it, and its manifestations across diverse cultural contexts.

## REFERENCES


Ahler, D. J., & Sood, G. (2018). The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics*, *80*(3), 964-981.

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of personality assessment*, *91*(4), 340-345.

Ballantyne, N. (2023). Recent work on intellectual humility: A philosopher's perspective. *The Journal of Positive Psychology*, *18*(2), 200-220.

Bastian, B., Zhang, A., & Moffat, K. (2015). The interaction of economic rewards and moral convictions in predicting attitudes toward resource use. *PLoS ONE, 10*(8), Article e0134863. https://doi.org/10.1371/journal.pone.0134863

Baumeister, R. F. (1999). Evil: Inside human violence and cruelty. Macmillan.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, *107*(2), 238.

Bowes, S. M., Blanchard, M. C., Costello, T. H., Abramowitz, A. I., & Lilienfeld, S. O. (2020). Intellectual humility and between-party animus: Implications for affective polarization in two community samples. *Journal of Research in Personality*, *88*, 103992.

Brandt, M. J., & Reyna, C. (2010). The role of prejudice and the need for closure in religious fundamentalism. *Personality and Social Psychology Bulletin*, *36*(5), 715-725.

Brandt, M. J., & Vallabha, S. (2022). Intraindividual Changes in Political Identity Strength (but not Direction) are Associated with Political Animosity in the United States and the Netherlands. *Personality and Social Psychology Bulletin*, 01461672231203471.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114*(28), 7313-7318.

Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. P. (2021). The primacy of morality in impression development: Theory, research, and future directions. In *Advances in experimental social psychology* (Vol. 64, pp. 187-262). Academic Press.

Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, *352*(6282), 220-224.

Burke, B. L., Martens, A., & Faucher, E. H. (2010). Two decades of terror management theory: A meta-analysis of mortality salience research. *Personality and Social Psychology Review*, *14*(2), 155-195

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of personality and social psychology*, *42*(1), 116

Cassese, E. C. (2021). Partisan dehumanization in American politics. *Political Behavior*, *43*(1), 29-50.

Clifford, S., & Rainey, C. (2024). Estimators for Topic-Sampling Designs. *Political Analysis*, *32*(4), 431-444.

Clifford S. (2019). How emotional frames moralize and polarize political attitudes. *Political Psychology*, 40(1), 75–91. https://doi.org/10.1111/pops.12507

Collier-Spruel, L., Hawkins, A., Jayawickreme, E., Fleeson, W., & Furr, R. M. (2019). Relativism or tolerance? Defining, assessing, connecting, and distinguishing two moral personality features with prominent roles in modern societies. *Journal of personality*, *87*(6), 1170-1188.

Comrey, A., & Lee, H. (1992). A first course in factor analysis. Hillsdale, NJ: Erlbaum.

Conrique, B. G. I. (2021). Different Values but Similar Backgrounds: How Relativism Influences Naïve Realism in Everyday Disagreements (Doctoral dissertation, University of Pittsburgh).

Corneille, O., & Gawronski, B. (2024). Self-reports are better measurement instruments than implicit measures. *Nature Reviews Psychology*, 1-12.

Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. *The evolution of morality*, 27-51.

Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current anthropology*, *60*(1), 47-69.

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated?. *Trends in cognitive sciences*, *24*(4), 267-269.

Davis, D. E., Rice, K., McElroy, S., DeBlaere, C., Choe, E., Van Tongeren, D. R., & Hook, J. N. (2016). Distinguishing intellectual humility and general humility. The Journal of Positive Psychology, 11(3), 215-224.

Darwin, C. (1872). The descent of man, and selection in relation to sex (Vol. 2). D. Appleton.

David, C (2023). The Chinese Room Argument, *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/sum2023/entries/chinese-room/>.

Delton, A. W., DeScioli, P., & Ryan, T. J. (2020). Moral obstinacy in political negotiations. Political Psychology, 41(1), 3-20.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *Plos one*, *18*(3), e0279720.

Drinkwater, K., Dagnall, N., Denovan, A., Parker, A., & Clough, P. (2018). Predictors and associates of problem–reaction–solution: statistical bias, emotion-based reasoning, and belief in the paranormal. *SAGE Open*, *8*(1), 2158244018762999.

Druckman, J. N., & Levy, J. (2022). Affective polarization in the American public. In *Handbook on politics and public opinion* (pp. 257-270). Edward Elgar Publishing.

Effron, D. A., & Miller, D. T. (2012). How the moralization of issues grants social legitimacy to act on one's attitudes. *Personality and Social Psychology Bulletin*, *38*(5), 690-701.

Enders, A. M., & Lupton, R. N. (2021). Value extremity contributes to affective polarization in the US. *Political Science Research and Methods*, *9*(4), 857-866.

Enders, A. M., & Armaly, M. T. (2019). The differential effects of actual and perceived polarization. *Political Behavior*, *41*, 815-839.

Faul, F., Earth fields, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analysis. *Behavior Research Methods*, *41*, 1149-1160

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., ... & Druckman, J. N. (2020). Political sectarianism in America. Science, 370(6516), 533-536.

Finkel, E., Landry, A., Hoyle, R. H., Druckman, J., & Van Bavel, J. J. (2024). Partisan Antipathy and the Erosion of Democratic Norms

Fiske, A. P., & Rai, T. S. (2014). Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships. Cambridge University Press.

Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of research in personality*, *56*, 82-92.

Fowers, B. J., Carroll, J. S., Leonhardt, N. D., & Cokelet, B. (2021). The emerging science of virtue. *Perspectives on Psychological Science*, *16*(1), 118-147.

Frankovic, K., Bialik, C., & Orth, T. (2023, March 30). Which issues Americans care about most and how they evaluate their current congressional leaders. YouGov. https://today.yougov.com/politics/articles/45495-which-issues-americans-care-about-most-poll?redirect_from=%2Ftopics%2Fpolitics%2Farticles-reports%2F2023%2F03%2F30%2Fwhich-issues-americans-care-about-most-poll

Frimer, J. A., Skitka, L. J., & Motyl, M. (2017). Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *Journal of Experimental Social Psychology*, *72*, 1-12.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in methods and practices in psychological science*, *2*(2), 156-168.

Furnham, A., & Thorne, J. D. (2013). Need for cognition: Its dimensionality and personality and intelligence correlates. *Journal of Individual Differences, 34*(4), 230–240. https://doi.org/10.1027/1614-0001/a000119

Garrett, K. N. (2019). Fired up by morality: The unique physiological response tied to moral conviction in politics. *Political Psychology*, *40*(3), 543-563..

Garrett, K. N., & Bankert, A. (2020). The moral roots of partisan division: How moral conviction heightens affective polarization. *British Journal of Political Science*, *50*(2), 621-640.

Goering, M., Espinoza, C. N., Mercier, A., Eason, E. K., Johnson, C. W., & Richter, C. G. (2024). Moral identity in relation to emotional well-being: a meta-analysis. *Frontiers in Psychology*, *15*, 1346732.

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, *106*(3), 1339-1366.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of personality and social psychology*, *106*(1), 148.

Gowans, C. (2021). Moral Relativism, *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2021/entries/moral-relativism/>.

Graham, M. H., & Svolik, M. W. (2020). Democracy in America? Partisanship, polarization, and the robustness of support for democracy in the United States. *American Political Science Review*, *114*(2), 392-409.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493-498.

Grubbs, J. B., Warmke, B., Tosi, J., James, A. S., & Campbell, W. K. (2019). Moral grandstanding in public discourse: Status-seeking motives as a potential explanatory mechanism in predicting conflict. *PloS one*, *14*(10), e0223749.

Grubbs, J. B., Warmke, B., Tosi, J., & James, A. S. (2020). Moral grandstanding and political polarization: A multi-study consideration. *Journal of Research in Personality*, *88*, 104009.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.

Haidt, J., & Kesebir, S. (2010). Morality. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 797–832). John Wiley & Sons, Inc.. https://doi.org/10.1002/9780470561119.socpsy002022

Halevy, N., Kreps, T. A., Weisel, O., & Goldenberg, A. (2015). Morality in intergroup conflict. *Current Opinion in Psychology*, *6*, 10-14.

Hare, J (2019). Religion and Morality, *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2019/entries/religion-morality/>.

Hill, E. D., Cohen, A. B., Terrell, H. K., & Nagoshi, C. T. (2010). The role of social cognition in the religious fundamentalism-prejudice relationship. *Journal for the Scientific Study of Religion*, *49*(4), 724-739.

Hobolt, S. B., Lawall, K., & Tilley, J. (2023). The polarizing effect of partisan echo chambers. *American Political Science Review*, 1-16.

Hooper, D., Coughlan, J. and Mullen, M. R. "Structural Equation Modelling: Guidelines for Determining Model Fit." The Electronic Journal of Business Research Methods Volume 6 Issue 1 2008, pp. 53 - 60, available online at www.ejbrm.com

Hoyle, R. H., Davisson, E. K., Diebels, K. J., & Leary, M. R. (2016). Holding specific views with humility: Conceptualization and measurement of specific intellectual humility. *Personality and Individual Differences*, *97*, 165-172.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, *15*(4), 309.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual review of political science*, *22*, 129-146.

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: a concise measure of the dark triad. *Psychological assessment*, *22*(2), 420.

Jung, J. H., & Clifford, S. (2024). Varieties of Values: Moral Values Are Uniquely Divisive. *American Political Science Review*, 1-17.

Kalmoe, N. P. & Mason, L. *Radical American Partisanship: Mapping Violent Hostility, Its Causes, and the Consequences for Democracy* (Univ. of Chicago Press, 2022).

Kingzette, J. et al. How affective polarization undermines support for democratic norms. *Public Opin. Q.* **85**, 663–677 (2021).

Kodapanakkal, R. I., Brandt, M. J., Kogler, C., & van Beest, I. (2022). Moral frames are persuasive and moralize attitudes; nonmoral frames are persuasive and de-moralize attitudes. *Psychological Science*, *33*(3), 433-449.

Kovacheff, C., Schwartz, S., Inbar, Y., & Feinberg, M. (2018). The problem with morality: Impeding progress and increasing divides. *Social Issues and Policy Review*, *12*(1), 218-257.

Kraaijeveld, S. R., & Jamrozik, E. (2022). Moralization and mismoralization in public health. *Medicine, Health Care and Philosophy*, *25*(4), 655-669.

Krumrei-Mancuso, E. J., & Newman, B. (2020). Intellectual humility in the sociopolitical domain. *Self and Identity*, *19*(8), 989-1016.

Kuhne, P. & Kamin, J. (n.d.) Social Cohesion Impact Measure (SCIM): A tool to measure the impact of bridging interventions. Retrieved from https://docs.google.com/document/d/1X2l379C7zYTH4EGOLilyMIig8QG7EdstVigvO7wGpFs/edit

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

Leary, M. R., Diebels, K. J., Davisson, E. K., Jongman-Sereno, K. P., Isherwood, J. C., Raimi, K. T., ... & Hoyle, R. H. (2017). Cognitive and interpersonal features of intellectual humility. *Personality and Social Psychology Bulletin*, *43*(6), 793-813.

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature human behaviour*, *4*(3), 279-286.

Lelkes, Y., & Westwood., S (2023). Global Political Pulse. https://americaspoliticalpulse.com/global

Levy, R. E. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review*, *111*(3), 831-870.

Lins de Holanda Coelho, G., HP Hanel, P., & J. Wolf, L. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, *27*(8), 1870-1885.

Liu, Q., & Nesbit, J. C. (2023). The relation between need for cognition and academic achievement: A meta-analysis. *Review of Educational Research*, 00346543231160474.

Luttrell, A., Petty, R. E., Briñol, P., & Wagner, B. C. (2016). Making it moral: Merely labeling an attitude as moral increases its strength. *Journal of Experimental Social Psychology*, *65*, 82-93.

MacAskill, M., Bykvist, K., & Ord, T. (2020). *Moral uncertainty* (p. 240). Oxford University Press.

Marie, A., Altay, S., & Strickland, B. (2023). Moralization and extremism robustly amplify myside sharing. *PNAS nexus*, *2*(4), pgad078.

Martherus, J. L., Martinez, A. G., Piff, P. K., & Theodoridis, A. G. (2021). Party animals? Extreme partisan polarization and dehumanization. *Political Behavior*, *43*, 517-540.

Mason, L. (2018). Ideologues without issues: The polarizing consequences of ideological identities. *Public Opinion Quarterly*, *82*(S1), 866-887.

McElroy-Heltzel, S. E., Davis, D. E., DeBlaere, C., Worthington Jr, E. L., & Hook, J. N. (2019). Embarrassment of riches in the measurement of humility: A critical review of 22 measures. The Journal of Positive Psychology, 14(3), 393-404.

McElroy, S. E., Rice, K. G., Davis, D. E., Hook, J. N., Hill, P. C., Worthington Jr, E. L., & Van Tongeren, D. R. (2014). Intellectual humility: Scale development and theoretical elaborations in the context of religious leadership. *Journal of Psychology and Theology*, *42*(1), 19-30

McCoy, J., & Somer, M. (2019). Toward a theory of pernicious polarization and how it harms democracies: Comparative evidence and possible remedies. *The Annals of the American Academy of Political and Social Science*, *681*(1), 234-271.

McCoy, J., Rahman, T., & Somer, M. (2018). Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. *American Behavioral Scientist, 62*(1), 16–42. https://doi.org/10.1177/0002764218759576

McGarry, P. P., Shteynberg, G., Hulsey, T. L., & Heim, A. S. (2023). The Great Divide: Neither Fairness Nor Kindness Eliminates Moral Derogation of People With Opposing Political Beliefs. *Social Psychological and Personality Science*, 19485506231194279.

McLaughlin, A. T., Van Tongeren, D. R., McElroy-Heltzel, S. E., Bowes, S. M., Rice, K. G., Hook, J. N., ... & Davis, D. E. (2023). Intellectual humility in the context of existential commitment. *The journal of positive psychology*, *18*(2), 289-303.

Mernyk, J. S., Pink, S. L., Druckman, J. N., & Willer, R. (2022). Correcting inaccurate metaperceptions reduces Americans' support for partisan violence. *Proceedings of the National Academy of Sciences*, *119*(16), e2116851119.

Miller, J. D., Back, M. D., Lynam, D. R., & Wright, A. G. (2021). Narcissism today: What we know and what we need to learn. *Current Directions in Psychological Science*, *30*(6), 519-525.

Moore-Berg, S. L., & Hameiri, B. (2024). Improving intergroup relations with meta-perception correction interventions. *Trends in Cognitive Sciences*

Moore-Berg, S. L., Ankori-Karlinsky, L.-O., Hameiri, B. & Bruneau, E. (2020).

Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proc. Natl Acad. Sci. USA* **117**, 14864–14872

Nicolas, G., Fiske, S. T., Koch, A., Imhoff, R., Unkelbach, C., Terache, J., Carrier, A., & Yzerbyt, V. (2022). Relational versus structural goals prioritize different social information. *Journal of Personality and Social Psychology, 122*(4), 659–682. https://doi.org/10.1037/pspi0000366

Olson, K., Camp, C., & Fuller, D. (1984). Curiosity and need for cognition. *Psychological reports*, *54*(1), 71-74

Owens, B. P., Yam, K. C., Bednar, J. S., Mao, J., & Hart, D. W. (2019). The impact of leader moral humility on follower moral self-efficacy and behavior. *Journal of Applied Psychology*, *104*(1), 146.

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual review of psychology*, *72*, 533-560.

Pasek, M. H., Ankori-Karlinsky, L. O., Levy-Vene, A., & Moore-Berg, S. L. (2022). Misperceptions about out-partisans' democratic values may erode democracy. *Scientific Reports*, *12*(1), 16284.

Peer, E., Rothschild, D., & Gordon, A. (2022). Data quality of platforms and panels for online behavioral research. Behav Res 54, 1643–1662 (2022).

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: the Epistemically-Related Emotion Scales. *Cognition and Emotion*, *31*(6), 1268-1276.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). Making sense of factor analysis. SAGE Publications, Inc., https://doi.org/10.4135/9781412984898

Pew Research Center (2021, May 4). 70% of U.S. social media users never or rarely post or share about political, social issues. Pew Research Center. https://www.pewresearch.org/short-reads/2021/05/04/70-of-u-s-social-media-users-never-or-rarely-post-or-share-about-political-social-issues/

Pew Research Center (2019, October 23). National politics on Twitter: Small share of U.S. adults produce majority of tweets. Pew Research Center. https://www.pewresearch.org/politics/2019/10/23/national-politics-on-twitter-small-share-of-u-s-adults-produce-majority-of-tweets/

Pinker, S. (2008). The Moral Instinct. New York Times Magazine (2008, January 13). https://www.nytimes.com/2008/01/13/magazine/13Psychology-t.html

Plante, T. G., & Boccaccini, M. T. (1997). The Santa Clara strength of religious faith questionnaire. *Pastoral Psychology*, *45*(5), 375-387.

Porter, T., Baldwin, C. R., Warren, M. T., Murray, E. D., Cotton Bronk, K., Forgeard, M. J., ... & Jayawickreme, E. (2022). Clarifying the content of intellectual humility: A systematic review and integrative framework. *Journal of Personality Assessment*, *104*(5), 573-585.

Prentice, M., Jayawickreme, E., Hawkins, A., Hartley, A., Furr, R. M., & Fleeson, W. (2019). Morality as a basic psychological need. *Social Psychological and Personality Science*, *10*(4), 449-460.

Pretus, C., Ray, J. L., Granot, Y., Cunningham, W. A., & Van Bavel, J. J. (2023). The psychology of hate: Moral concerns differentiate hate from dislike. European Journal of Social Psychology, 53(2), 336-353.

Priem, R. L., Wenzel, M., & Koch, J. (2018). Demand-side strategy and business models: Putting value creation for consumers center stage. *Long range planning*, *51*(1), 22-31.

Puryear, C., Kubin, E., Schein, C., Bigman, Y., & Gray, K. (2022). Bridging political divides by correcting the basic morality bias. Retrieved from: https://osf.io/preprints/psyarxiv/fk8g6

Reicher, S., Haslam, S. A., & Rath, R. (2008). Making a virtue of evil: A five-step social identity model of the development of collective hate. *Social and Personality Psychology Compass*, *2*(3), 1313-1344.

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 27, 151-161.

Rozin, P. (2006). Domain denigration and process preference in academic psychology. *Perspectives on Psychological Science*, *1*(4), 365-376.

Ruberton, P. M., Kruse, E., & Lyubomirsky, S. (2016). 18Boosting State Humility via Gratitude, Self-Affirmation, and Awe: Theoretical and Empirical Perspectives. In *Handbook of humility* (pp. 276-289). Routledge.

Ruggeri, K., Većkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., ... & Folke, T. (2021). The general fault in our fault lines. *Nature Human Behaviour*, *5*(10), 1369-1380.

Ryan, T. J. (2014). Reconsidering moral issues in politics. *The Journal of Politics*, *76*(2), 380-397.

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

Salgado, J. F. (2017). Bandwidth-fidelity dilemma. *Encyclopedia of personality and individual differences*, 1-4.

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, *22*(1), 32-70.

Schmidt, K., Jones, A. L., Szabelska, A., Ebersole, C. R., Hawkins, C. B., Graham, J., & Nosek, B. A. (2023, October 17). The Ideology 2.0 Study and Dataset. Retrieved from osf.io/2483h

Schroeders, U., Wilhelm, O., & Olaru, G. (2016). Meta-heuristics in short scale construction: Ant colony optimization and genetic algorithm. *PloS one*, *11*(11), e0167110.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. psychology press.

Sedikides, C., Meek, R., Alicke, M. D., & Taylor, S. (2014). Behind bars but above the bar: Prisoners consider themselves more prosocial than non-prisoners. *British Journal of Social Psychology*, *53*(2), 396-403.

Sibley, C. G., Osborne, D., & Duckitt, J. (2012). Personality and political orientation: Meta-analysis and test of a Threat-Constraint Model. *Journal of Research in Personality*, *46*(6), 664-677.

Simonovits, G., McCoy, J., & Littvay, L. (2022). Democratic hypocrisy and out-group threat: explaining citizen support for democratic erosion. *The Journal of Politics*, *84*(3), 1806-1811.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68*, 69-81.

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.

Skitka, L. J., Hanson, B. E., Morgan, G. S., & Wisneski, D. C. (2021). The psychology of moral conviction. Annual Review of Psychology, 72, 347-366.

Skitka, L. J., & Mullen, E. (2002). The dark side of moral conviction. Analyses of Social Issues and Public Policy, 2(1), 35-41.

Smith, I. H., & Kouchaki, M. (2018). Moral humility: In life and at work. *Research in Organizational Behavior*, *38*, 77-94.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159-171.

Stanley, M. L., Bedrov, A., Cabeza, R., & De Brigard, F. (2020). The centrality of remembered moral and immoral actions in constructing personal identity. *Memory*, *28*(2), 278-284.

Tangney, J. P. (2009). Humility. In S. J. Lopez & C. R. Snyder (Eds.), *Oxford handbook of positive psychology* (2nd ed., pp. 483–490). Oxford University Press.

Tannenbaum, M. B., Hepler, J., Zimmerman, R. S., Saul, L., Jacobs, S., Wilson, K., & Albarracín, D. (2015). Appealing to fear: A meta-analysis of fear appeal effectiveness and theories. *Psychological bulletin*, *141*(6), 1178.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis.

Tersman, Folk (2022). Moral Disagreement, *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2022/entries/disagreement-moral/>.

Tosi, J., & Warmke, B. (2020). *Grandstanding: The use and abuse of moral talk*. Oxford University Press, USA.

Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual review of psychology*, *64*, 231-255.

Tracy, J. L., Cheng, J. T., Robins, R. W., & Trzesniewski, K. H. (2009). Authentic and hubristic pride: The affective core of self-esteem and narcissism. *Self and identity*, *8*(2-3), 196-213.

Vallabha, S. & Brandt, M. J. (2024, April 5). Moral Humility Reduces Political Divisions. https://doi.org/10.31234/osf.io/5r6fw

Van Bavel, J. J., Packer, D. J., Haas, I. J., & Cunningham, W. A. (2012). The importance of moral construal: Moral versus non-moral construal elicits faster, more extreme, universal evaluations of the same actions. PloS one, 7(11), e48693.

Van Bavel, J. J., Robertson, C. E., Del Rosario, K., Rasmussen, J., & Rathje, S. (2024). Social media and morality. *Annual review of psychology*, *75*, 311-340.

Van Tongeren, D. R., Stafford, J., Hook, J. N., Green, J. D., Davis, D. E., & Johnson, K. A. (2016). Humility attenuates negative attitudes and behaviors toward religious out-group members. *The Journal of Positive Psychology*, *11*(2), 199-208.

Van Tongeren, D. R., Davis, D. E., Hook, J. N., & Witvliet, C. vanOyen. (2019). Humility. Current Directions in Psychological Science, 28(5), 463-468. https://doi.org/10.1177/0963721419850153

Van Tongeren, D. R., Ng, V., Hickman, L., & Tay, L. (2023). Behavioral measures of humility: Part 2. Conceptual mapping and charting ways forward. *The Journal of Positive Psychology*, *18*(5), 722-732.

Van Zomeren, M., Postmes, T., & Spears, R. (2012). On conviction's collective consequences: Integrating moral conviction with the social identity model of collective action. *British Journal of Social Psychology*, *51*(1), 52-71.

Voelkel, J. G., Chu, J., Stagnaro, M. N., Mernyk, J. S., Redekopp, C., Pink, S. L., ... & Willer, R. (2023). Interventions reducing affective polarization do not necessarily improve anti-democratic attitudes. *Nature human behaviour*, *7*(1), 55-64.

Voelkel, J. G., Stagnaro, M., Chu, J., Pink, S. L., Mernyk, J. S., Redekopp, C., … Willer, R. (2023, March 20). Megastudy identifying effective interventions to strengthen Americans' democratic attitudes. https://doi.org/10.31219/osf.io/y79u5.

Wilson, T. D., Aronson, E., & Carlsmith, K. (2010). The art of laboratory experimentation. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology* (5th ed., Vol. 1, pp. 51-81). Hoboken, NJ: Wiley.

Williams, E. G. (2015). The possibility of an ongoing moral catastrophe. *Ethical Theory and Moral Practice*, *18*, 971-982.

Wilson, T. D., Aronson, E., & Carlsmith, K. (2010). The art of laboratory experimentation. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology* (5th ed., Vol. 1, pp. 51-81). Hoboken, NJ: Wiley.

Wright, J. C., Nadelhoffer, T., Perini, T., Langville, A., Echols, M., & Venezia, K. (2017). The psychological significance of humility. The Journal of Positive Psychology, 12(1), 3-12.

Wright, J. C., & Pölzler, T. (2022). Should morality be abolished? An empirical challenge to the argument from intolerance. *Philosophical psychology*, *35*(3), 350-385.

Yoder, K. J., & Decety, J. (2014). Spatiotemporal neural dynamics of moral judgment: A high-density ERP study. *Neuropsychologia*, *60*, 39-45.

Zimbardo, Philip G. "A situationist perspective on the psychology of evil: Understanding how good people are transformed into perpetrators." *The social psychology of good and evil* (2004): 21-50.

Zmigrod, L., Rentfrow, P. J., Zmigrod, S., & Robbins, T. W. (2019). Cognitive flexibility and religious disbelief. *Psychological research*, *83*(8), 1749-1759.