

THE UNIVERSITY OF CHICAGO

THE ROLE OF MORAL CONVICTION IN NEURAL, ATTENTIONAL, AND
METACOGNITIVE PROCESSES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY
QIONGWEN CAO

CHICAGO, ILLINOIS

AUGUST 2026

TABLE OF CONTENTS

LIST OF FIGURES	iii
LIST OF TABLES.....	v
ACKNOWLEDGEMENTS.....	vi
ABSTRACT.....	vii
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 MORAL CONVICTION AND METACOGNITION SHAPE NEURAL RESPONSE DURING SOCIOPOLITICAL DECISION-MAKING.....	8
<i>Introduction</i>	9
<i>Methods</i>	15
<i>Results</i>	29
<i>Discussion</i>	41
<i>Conclusion</i>	47
<i>Appendix A</i>	49
CHAPTER 3 ATTENTION AND COGNITIVE MECHANISMS IN SOCIOPOLITICAL DECISION-MAKING ABOUT MORALLY CONVICTED ISSUES	51
<i>Introduction</i>	52
<i>Methods</i>	58
<i>Results</i>	63
<i>Discussion</i>	68
CHAPTER 4 MORAL CONVICTION AFFECTS FACTUAL JUDGMENTS AND METACOGNITIVE ABILITIES	71
<i>Introduction</i>	72
<i>Methods</i>	80
<i>Results</i>	91
<i>Discussion</i>	96
<i>Appendix B</i>	102
CHAPTER 5 GENERAL DISCUSSION	103
<i>Appendix C</i>	109
<i>Appendix D</i>	127
REFERENCES	146

LIST OF FIGURES

Figure 2.1: Mean moral conviction ratings for each sociopolitical issue from the initial online survey. Error bars represent one standard deviation from the mean.	17
Figure 2.2: Example of stimulus sequence in a trial during fMRI scanning.	19
Figure 2.3: Effect of religiosity on moral conviction. The shades illustrate 95% confidence intervals.	30
Figure 2.4: Relationship between the probability of choosing to support protesters on the left and the relative support rating of the issue on the left. The shade illustrates the 95% confidence interval.	31
Figure 2.5: Effects of maximum moral conviction, support difference, and protesters' position on response time. The shades illustrate 95% confidence intervals.	32
Figure 2.6: Parametric effects of support of sociopolitical issues during decision-making.	33
Figure 2.7: Parametric effects of moral conviction of sociopolitical issues during decision-making.	34
Figure 2.8: A) IPFC seed region and B) gPPI connectivity higher maximum moral conviction.	35
Figure 2.9: A) Relationship between probability of choosing to support protesters on the left-hand side of the screen and the relative support rating of the issue from the pre-scan survey, moderated by the strength of an individual's parametric effect of moral conviction on brain response in regions shown in Figure 2.7. B) Relationship between decision time and maximum moral conviction on each trial, moderated by the individual's parametric effect of moral conviction on brain response in regions shown in Figure 2.7. Shadings indicate 95% confidence intervals.	38
Figure 2.10: Negative correlations between meta-d' and the parametric effects of moral conviction A) across all responsive regions and B) in cortical valuation regions (VS and vmPFC).	40
Figure 2.11: Effects of maximum moral conviction, support difference, and maximum support on response time. Note that only maximum moral conviction and maximum support had a significant impact on response time, whereas support difference did not. The shades illustrate 95% confidence intervals.	50
Figure 3.1: Fixed-effect coefficient plot from the model predicting the relative left dwell time proportion with 95% confidence intervals.	64
Figure 3.2: Effects of support difference, moral difference, and moral misalignment on total dwell count in a trial.	66

Figure 3.3: Illustrative DDM trajectories showing separate effects on drift rate (v) and boundary separation (a).	68
Figure 4.1: Summary of the 2x2 (high/low moral conviction x consistent/inconsistent) conditions of the novel semantic knowledge metacognition task. A, an example of the semantic knowledge task (<i>high moral conviction, (in)consistent</i> if participant opposes(supports) capital punishment), correct answer is option <i>a</i> , not in favor of capital punishment. B, an example of the semantic knowledge task, (<i>high moral conviction, (in)consistent</i> if participant supports(opposes) capital punishment), correct answer is option <i>a</i> , more in favor of capital punishment compared to option <i>b</i> . C, an example of the semantic knowledge task (<i>low moral conviction, (in)consistent</i> if participant prefers binge watching (spaced viewing), correct answer is option <i>a</i> , in favor of binge watching. D, an example of the semantic knowledge task, (<i>low moral conviction, (in)consistent</i> if participant prefers spaced viewing (binge watching), the correct answer is option <i>b</i> , not in favor of binge watching.....	83
Figure 4.2: Moral conviction by topics. Ratings are shown for a subset of 140 participants (including 126 participants from the second batch), because minor changes were made to some topics and questions across batches.	86
Figure 4.3: Effects of moral conviction and fact-view consistency on accuracy. The shades illustrate 95% confidence intervals.	92
Figure 4.4: Effects of moral conviction and accuracy on confidence. The shades illustrate 95% confidence intervals.	93
Figure 4.5: Effects of fact-view consistency, accuracy and moral conviction on confidence. The shades illustrate 95% confidence intervals.	94
Figure 4.6: Correlation matrix of metacognitive efficiency in the perceptual, memory, low moral conviction, and high moral conviction domains. Heatmap colors represent strength of association.	95
Figure 4.7: Effects of fact-view consistency, moral conviction, and party alignment on factual judgment correctness. The shades illustrate 95% confidence intervals.....	102

LIST OF TABLES

Table 2.1: Brain regions showing significant parametric effect of mean support ratings across the two issues in a given trial.	33
Table 2.2: Brain areas showing significant parametric effects of maximum moral conviction rating on hemodynamic response.	35
Table 2.3: Brain regions that showed significant increases in functional connectivity with IPFC during decisions with higher maximum moral conviction.	36
Table 3.1: Drift diffusion model estimates for support difference, moral misalignment, and gaze bias. The table summarizes key fixed effects from the DDM and the decision parameters with which they were associated.	67

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor, Jean Decety, for his guidance, support, and trust in my ability to become a good and independent researcher. His enthusiasm and passion for research have always been contagious and deeply inspiring. He has taught me to learn from every experience – successes, failures, and everything in between. I am truly grateful. I would also like to thank my committee members, Akram Bakkour, Wilma Bainbridge and Tomas Ståhl, who have been so encouraging and have offered thought-provoking questions, brilliant comments, and insightful suggestions. Thank you for going out of your way to help me become a stronger and wiser researcher.

In addition to my committee, I want to express my gratitude to Keith Yoder and Michael Cohen, two important collaborators who have shaped this dissertation in meaningful ways. I am grateful to Keith for introducing me to fMRI and moral conviction. I am also thankful to Michael for always being so kind and generous in sharing his knowledge and for discussing with me everything from big-picture theoretical questions to the finer details of the analyses. It has truly been my pleasure to work with them. Throughout my PhD, I have also been fortunate to receive support and guidance from many scholars and collaborators, some of them across the country or around the world. Specifically, I want to thank Fan Yang, Yuan Chang Leong, and Fang Cui for all the feedback on my work and the collaborative work we have done together. I also owe a special thank-you to John Monterosso, who introduced me to the world of research during the early years in college and, since then, has been an extraordinary mentor and friend.

Lastly, I want to thank my parents and friends. Your support has carried me through the most difficult moments of this journey and made the joyful moments even more meaningful. I could not have done this without you. I love you all.

ABSTRACT

Moral conviction refers to the extent to which attitudes are experienced as moral commitments, rooted in one's core values and concerns about right and wrong. It can promote civic engagement and collective action, but also lead to dogmatism, polarization, and even violence toward those who disagree. Yet the mechanisms through which moral conviction affects behaviors are less understood. This dissertation addresses this gap by examining the role that moral conviction plays in neural, attentional, and cognitive processes during sociopolitical decisions and metacognitive judgments. Across studies, results show that: 1) moral conviction engages both regions in the salience network and lateral prefrontal cortex, supporting the theoretical account that moral conviction consists of both cognitive and affective dimensions; 2) moral conviction may have distinct effects depending on its relation to participants' prior beliefs: overall moral conviction, agnostic to belief congruence, speeds up decisions and attracts attention, whereas moral misalignment, the extent to which a belief-incongruent position is moralized, slows responses, increases dwell count and promotes decision caution; and 3) moral conviction boosts confidence without improving accuracy in factual judgments, with confidence-accuracy coupling being especially poor when facts conflict with individuals' strongly moralized beliefs. Together, this work illuminates the attentional and neural mechanisms by which moral conviction shapes sociopolitical decisions. It also demonstrates that moral conviction biases both factual judgments and metacognitive monitoring, offering psychological explanations for broader societal dynamics, such as factual disagreement and polarization.

CHAPTER 1

INTRODUCTION

From an evolutionary perspective, morality has evolved throughout human history as an adaptation that fosters cooperation (Tomasello, 2016). Humans not only interdepend on and collaborate with particular others to achieve mutual benefit, but they also scale up these interpersonal collaborative skills to group life, establishing impersonal, agent-neutral, and objective social norms (Tomasello & Vaish, 2013). These norms encompass a variety of rules that people follow, which promote the predictability of others' behaviors, maintain order in society, and help regulate large-scale interaction and cooperation. While some norms are specific to particular times, places, and whether approved by authorities, and are therefore categorized as conventions, others become moralized, where originally neutral attitudes or preferences are converted into values perceived as universal and independent of authority (Nucci & Turiel, 1978; Rozin, 1999). Studying moral attitudes toward smoking cigarettes across generations, Rozin and Singh (1999) found there had been an increase in health and moral concerns about smoking over the preceding 20-40 years. In the reverse direction, moral values can gradually be detached from a substance or activity, such as the use of certain drugs and sexual orientation. These studies suggest that, while both conventions and morality may have originated in part to facilitate the functioning of large-scale societies, they remain conceptually distinct at a given historical moment. From a longitudinal perspective, however, the extent to which a concept, object, or opinion is perceived as morally significant is fluid and evolves alongside broader societal changes.

As the field of moral psychology has grown, researchers have used and built upon Rozin's definition of moralization to study moral change. Rhee and colleagues (2019)

summarized that the moralization process has been examined in two forms: moral recognition and moral amplification. Moral recognition refers to the process of attaching or detaching moral significance to a given attitude, concept, or object. Moral emotions (e.g., disgust) and moral piggybacking have been suggested to be two promising mechanisms for moral recognition (Feinberg et al., 2019). Specifically, the more strongly individuals experience emotions such as guilt and disgust in response to an issue, the more likely they are to undergo moralization. At the same time, the more individuals engage in moral piggybacking, the process of connecting a particular issue at hand (e.g., eating animals) to their preexisting fundamental moral principles (e.g., one should avoid causing harm to others), the more they moralize the issue at a later time point. Consistent with these two mechanisms, research has found that individuals who moralize behaviors such as vegetarianism associate more disgust with those behaviors and offer more reasons in support of their stance (Rozin et al., 1997). On the other hand, moral amplification is the increase in the strength and extremity of moralization after moral relevance has already been attached. Some evidence has shown that emotions tied to the relevant attitude, framing behaviors in emotional or harm-based terms, and perceiving a given action as descriptive norm can all contribute to moral amplification (Brandt et al., 2015; Clifford, 2019; Rottman et al., 2017; Lindström et al., 2018).

Such cognitive (moral piggybacking, framing behaviors as harm-based terms) and affective (incidental emotions induction, emotional framing of a behavior) processes during moralization seem to converge with the literature that examines moral conviction. Skitka and colleagues (2005, see also Skitka & Mullen, 2002) developed and formalized the research construct of moral conviction, and defined it as “a strong and absolute belief that something is right or wrong, moral or immoral” (Skitka et al., 2005, p. 896). In other contexts, moral

conviction is operationalized using questions like “To what extent do you feel your feelings about X deeply connected to your fundamental beliefs about ‘right’ and ‘wrong?’” (Skitka et al., 2009; Skitka & Wisneski, 2011). Therefore, “moral conviction” can be the noun for the morally convicted beliefs themselves and the extent to which an individual moralizes a belief. In this dissertation, moral conviction refers to the latter in most cases but can sometimes refer to the former depending on context.

That said, studies investigating the functional architecture of moral conviction have proposed that it comprises cognitive and affective dimensions (Wright, 2012; Wright et al., 2008). The cognitive dimension is people’s categorization of a belief as moral or non-moral, and the affective dimension is the emotional intensity attached to the belief. This theoretical account is consistent with the moralization process in which a neutral belief becomes a moral conviction as people cognitively recognize the moral importance of an issue and attach emotions to it. While some behavioral studies have examined the two dimensions of moral conviction and their effects on various behaviors, few have investigated the mechanisms through which moral conviction exerts these effects. The first aim of this dissertation is to illustrate the psychological processes individuals experience when engaging with morally convicted attitudes. To do so, it adopts multiple methods. First, it uses behavioral analyses to examine how moral conviction affects sociopolitical decisions and their speed. Second, it uses fMRI to characterize neural activity associated with decisions about issues held with moral conviction. Lastly, it uses eye-tracking and computational modeling of decision-making (drift-diffusion model, DDM) to investigate how moral conviction affects visual attention and decision processes. Through this multi-level approach, this dissertation offers a clearer view of the psychological and neural processes

through which moral conviction affects decisions, rather than merely documenting its behavioral consequences.

The second focus of this dissertation is to elucidate the relationship between moral conviction and metacognition, an area that has yielded intriguing findings but remains relatively underexplored. In Skitka and colleagues' (2021) review, the authors emphasized the role of moral conviction as a meta-perception of "the degree to which a specific attitude reflects their moral concerns". It can therefore be related to the broadly defined metacognitive processes that involve monitoring and evaluating one's own mental states. More precisely, when an attitude is strongly morally convicted, people usually have the metacognitive understanding that such attitudes and beliefs are objective and universal. By contrast, in a narrower sense and in the more formal metacognition literature, metacognition refers more specifically to metacognitive evaluation of one's cognitive performance, which can be indicated by metrics like confidence-accuracy correspondence and metacognitive sensitivity, i.e., how well confidence discriminates between correct and incorrect responses (Yeung & Summerfield, 2012; Maniscalco & Lau, 2012). Although moral conviction and metacognition are far from identical, they share a second-order feature that both involve evaluations of one's own mental representations. This conceptual overlap raises the possibility that they may interact and have behavioral consequences.

There are two likely ways that moral conviction and metacognition interact. First, as studied in previous work, metacognition can be viewed as an individual-difference variable that affects one's belief systems or moderates the impact of moral conviction on other behaviors. For example, Rollwage and colleagues (2018) demonstrated that people who have worse metacognitive sensitivity also hold more radical political beliefs. In line with this finding, Yoder and Decety (Yoder & Decety, 2022) found that among people who have poorer metacognitive

sensitivity, moral conviction reduced social conformity to a greater extent, suggesting metacognition can modulate moral conviction's impact on behavior. In both studies, metacognitive sensitivity was measured in a separate low-level perceptual discrimination task where participants made perceptual judgments and rated their confidence in those judgments. It is striking that such seemingly unrelated perceptual metacognitive abilities are related to moral beliefs and behaviors. This dissertation further extends this line of research in examining how individual differences in perceptual metacognition may moderate people's neural response to moral conviction.

Meanwhile, another less studied way in which moral conviction and metacognition can interact is that moral conviction may have a direct impact on how well individuals discriminate between correct and incorrect judgments. Ryan (2019) has found that moral conviction is associated with thinking about political issues in absolutist terms and a lower likelihood of being persuaded by consequence-based disagreeing arguments. One possible outcome of such resistance to factual counterarguments is that individuals may become less accurate in judging whether what they believe about highly moralized topics is actually factually correct. This dissertation develops a new metacognition task to examine this possibility. Specifically, classic cognitive psychology research on metacognition has focused primarily on perception, memory, semantic knowledge, and decision-making. To examine whether moral conviction affects metacognition, the most suitable approach is to develop a task within the semantic knowledge framework. In a typical perceptual or memory metacognition task, individuals see some stimuli and make a series of 2-alternative forced-choice judgments about the stimuli they just saw, followed by a retrospective confidence judgment for their response. In a typical semantic knowledge metacognition task, however, individuals make judgments about general information

rather than about the stimuli they engage with in the experiment. It is reasonable to develop general information questions concerning topics that vary in moral conviction.

In sum, beyond examining the psychological processes involved when individuals engage attitudes held with moral conviction, this dissertation addresses the relationship between moral conviction and metacognition in two complementary ways: first, by conceptualizing metacognition as an individual difference variable and testing whether metacognitive sensitivity is related to variation in neural responses to moral conviction; and second, by conceptualizing metacognition as a context-sensitive process and testing whether the moral conviction level of the content individuals engage with influences factual judgments, confidence and metacognitive ability.

Overall, the research presented in this dissertation provides novel perspectives on moral conviction and metacognition. It has three major theoretical contributions. Initially, it contributes to shifting the literature from asking only what moral conviction predicts to asking how it shapes ongoing psychological processing, providing insight into the neural mechanisms of moral conviction (Chapter 2) and the visual attentional processes through which moral conviction may influence decision-making (Chapter 3). Moreover, it bridges moral psychology and metacognition research by examining how metacognition and moral conviction interact to shape neural responses (Chapter 2) and how morally convicted content influences metacognitive processes (Chapter 4). Finally, it expands research on metacognition into morally and politically consequential contexts by moving beyond the field's traditional focus on perception and memory to examine contested, value-laden, sociopolitical domains, while introducing a new task for studying this empirically (Chapter 4). In doing so, this research contributes to a deeper theoretical understanding of moral conviction. This mechanistic work may provide a foundation

for future work aimed at identifying ways to amplify moral conviction's positive role in civic engagement and collective action while mitigating its negative consequences, such as polarization and intolerance.

CHAPTER 2

MORAL CONVICTION AND METACOGNITION SHAPE NEURAL RESPONSE DURING SOCIOPOLITICAL DECISION-MAKING¹

The extent to which a belief is rooted in one's sense of morality has significant societal implications. While moral conviction can inspire positive collective action, it can also prompt dogmatism, intolerance, and societal divisions. Research in social psychology has documented the functional characteristics of moral conviction and shows that poor metacognition exacerbates its negative outcomes. However, the cognitive and neural mechanisms underlying moral conviction, their relationship with metacognition, and how moral conviction is integrated into the valuation and decision-making process, remain unclear. This chapter investigated these neurocognitive processes during decision-making on sociopolitical issues varying in moral conviction. Participants (N = 44) underwent fMRI scanning while deciding, on each trial, which of two groups of political protesters they supported more. As predicted, stronger moral conviction was associated with faster decision times. Hemodynamic responses in the anterior insula (aINS), anterior cingulate cortex (ACC), and lateral prefrontal cortex (LPFC) were elevated during decisions with higher moral conviction, supporting the emotional and cognitive dimensions of moral conviction. Functional connectivity between LPFC and vmPFC was greater on trials higher in moral conviction, elucidating mechanisms through which moral conviction is

¹ A version of Chapter 2 has been published as: Cao, Q., Cohen, M. S., Bakkour, A., Leong, Y. C., & Decety, J. (2025). Moral conviction interacts with metacognitive ability in modulating neural activity during sociopolitical decision-making. *Cognitive, Affective, & Behavioral Neuroscience*, 25(2), 291–310.

incorporated into valuation. Average support for the two displayed groups of protesters was positively associated with brain activity in regions involved in valuation, particularly vmPFC and amygdala. Metacognitive sensitivity, the ability to discriminate one's correct from incorrect judgments, measured in a perceptual task, negatively correlated with parametric effects of moral conviction in the brain, providing new evidence that metacognition modulates responses to morally convicted issues.

Introduction

Morally convicted attitudes pertain to beliefs regarding what is fundamentally right and wrong. These attitudes reflect core moral values, which are perceived as culturally universal absolutes, stable over time, and particularly resistant to authority (Skitka et al., 2021; Luttrell & Togans, 2021). Moral convictions can inspire benevolent forms of collective action, such as the American civil rights movement, but they can also incite dogmatism, division, and authoritarianism (Decety, 2024; Garrett & Bankert, 2020). Other harmful consequences include aggressive attitudes, hate speech, justification of prejudice, vigilantism, and physical violence against people or groups who share different values or practices (Workman et al., 2020; Wright & Pölzler, 2022; Yoder & Decety, 2022), all of which can have a corrosive effect on democracy (Finkel et al., 2020).

Moral conviction incorporates cognitive and affective dimensions (Wright et al., 2008). The cognitive dimension refers to the distinction between beliefs that are or are not moralized and accounts for the fact that the former are treated almost as objective and universal truth (Goodwin & Darley, 2008). The affective dimension, in contrast, reflects the emotional intensity associated with these beliefs. Research in social psychology and political science suggests that

when people moralize their opinions or attitudes, they are less likely to take in new information or consider arguments based on cost/benefit analysis (Ryan, 2019). This heightened moral conviction not only affects individuals' cognition but also scales up and leads to actions that influence others. For example, moral conviction has been linked to greater chances of sharing politically congruent partisan news, regardless of its veracity (Marie et al., 2023). Another study analyzed more than 500,000 tweets and demonstrated that messages increase their reach by 20% with each additional moral-emotional word (Brady et al., 2017). People are also more willing to engage in normative and nonnormative collective action when confronted with attitudes that strongly, rather than weakly, violate their moral beliefs (Pauls et al., 2022; Thomas et al., 2019). Meanwhile, encountering others holding opposing beliefs elicits negative emotions, particularly if these beliefs are perceived as moral obligations (Ryan, 2014; Zaal et al., 2017). Reading about more morally convicted political information predicts stronger physiological arousal measured by skin conductance, further supporting the contention that moral conviction evokes strong emotions (Garrett, 2019). Overall, these behavioral findings support the idea that morally convicted beliefs, compared to beliefs held without moral conviction, are cognitively perceived as more absolute and universal, and emotionally as higher in salience, motivating people to exert considerable effort to persuade others and achieve their moral objectives.

As moral conviction both elicits strong emotions and leads to the cognitive characterization of attitudes as largely absolute, judgments and behaviors about highly morally convicted issues may be faster and more consistent across time and scenarios. Indeed, research has found that moral evaluations of actions are faster than pragmatic evaluations, and individuals who perceived a greater moral basis of their attitudes showed a stronger correlation between their attitudes and behavioral intentions (Luttrell et al., 2016; Van Bavel et al., 2012). It remains

underexplored; however, how moral conviction affects sociopolitical decisions. The current study introduces a novel task where participants choose between two groups of protesters and decide which group they support more. This design enables us to investigate how moral conviction levels influence decision time and consistency between attitudes and decisions.

Prior research on the cognitive neuroscience of morality suggests a pivotal role for the lateral prefrontal cortex (LPFC) in the flexible implementation of social norms and the pursuit of moral goals (Carlson & Crockett, 2018; Yoder & Decety, 2018). Proper functioning of the dorsolateral prefrontal cortex (dlPFC) seems necessary for individuals to act cooperatively, synthesize the intentions of wrongdoers and the perceived harm to the victim, and determine punishments for moral and norm violations (Decety & Cowell, 2018; Krueger & Hoffman, 2016; Soutschek et al., 2015). Meanwhile, behavioral research has found that categorizing a belief as moral, relying on a cognitive judgment, leads to less tolerance towards those holding divergent views, while holding a belief with greater emotional intensity alone does not (Wright et al., 2008). Thus, punitive decisions studied in previous cognitive neuroscience studies can be understood as extensions of this intolerance and as consequences arising from the cognitive recognition of goals as moral. Other cognitive neuroscience studies have documented the implication of the anterior insula (aINS), anterior cingulate cortex (ACC), and amygdala in assessing the outcomes of interpersonal actions, especially harm, in a moral context (FeldmanHall & Mobbs, 2015; Hesse et al., 2016). The salience network, a suite of interconnected cortical and subcortical regions including the ACC, aINS, amygdala, ventral striatum, periaqueductal gray and ventral tegmental area, has a crucial role in detecting behaviorally relevant information and coordinating neural resources (Uddin, 2015). Specifically, the amygdala plays an essential role in directing attention to motivationally relevant and

emotionally arousing stimuli regardless of valence (Cunningham & Brosch, 2012; Lin et al., 2020; Wang et al., 2017). The aINS, a key region in the salience network, serves a related function in tracking overall emotional appraisals of moral situations (Hutcherson et al., 2015; Shenhav & Greene, 2014). Another major node in the salience network, the ACC, monitors the strength of emotional reactions to specific events (Seamans & Floresco, 2022). Together, it is reasonable to predict that the IPFC, amygdala, aINS, and ACC might be among the brain regions in which activity during decision-making tracks the levels of moral conviction.

Functional neuroimaging studies have also consistently identified a set of interconnected regions underlying moral decision-making. This circuit includes the ventromedial prefrontal cortex (vmPFC), the ventral striatum (VS), the IPFC, and aINS (Qu et al., 2022; Yoder & Decety, 2018). Consistent with the common-currency hypothesis, some studies have demonstrated that the valuation system, including the vmPFC and VS, tracks the subjective value of voluntary donations and the appropriateness ratings of sociopolitical violence (Clithero et al., 2011; Hare et al., 2010; Workman et al., 2020). Other studies have shown that distinct brain regions, including the temporoparietal junction (TPJ), IPFC, and insula, track moral values (Crockett et al., 2017; Qu et al., 2020; Ugazio et al., 2022). However, most past research has used heterogeneous tasks whose moral relevance has not been systematically measured. Hence, it is difficult to determine whether activity in the reported regions is specifically due to the moral significance of the issues being tested or because of related features such as empathy and mentalizing. In the present study, while fMRI data were being recorded, participants made decisions on which of two protester groups they supported more. Prior to scanning, the support and moral conviction levels of each issue had been measured. Thus, it was possible to quantitatively link these measures to neural mechanisms during the decision-making process and

relate the strength of these neural responses to behavioral measures such as decision time and attitude-decision consistency.

Cognitive inflexibility plays a role in moral conviction and is generally seen as an outcome of the moralization process (Skitka et al., 2005). Relatedly, evidence from multiple studies suggests that strong and dogmatic opinions are associated with a cognitive style that includes low metacognitive sensitivity (Rollwage et al., 2018; Yoder & Decety, 2022; Zmigrod et al., 2020). Metacognitive sensitivity reflects the degree to which confidence differentiates between correct and incorrect responses. Notably, the radical and dogmatic opinions examined are sociopolitical in nature, whereas metacognitive sensitivity is measured by an unrelated perceptual task. These findings are consistent with work showing that metacognition is at least partly domain-general (Mazancieux et al., 2020). While the precise mechanistic commonalities between perceptual metacognition and the formation of sociopolitical beliefs are not fully understood, the domain-generality of metacognition helps us begin to make sense of this perhaps-surprising relationship that has emerged across several distinct studies. Note that metacognitive sensitivity should not be confused with metacognitive bias, which is the overall level of overconfidence or underconfidence (Fleming & Lau, 2014). Low metacognitive sensitivity refers to a reduced capacity to distinguish between an individual's own correct and incorrect decisions; in such cases, after answering a question, individuals' confidence judgments are not well-aligned with their actual performance. Importantly, in a study in which both metacognitive sensitivity and overconfidence were measured, only metacognitive sensitivity showed a significant negative association with dogmatism while overconfidence did not (Rollwage et al., 2018). Furthermore, in that study, in individuals holding politically radical beliefs, poor metacognitive sensitivity is driven by unreasonably high confidence, particularly in

incorrect decisions, compared to moderates. Another previous study found that individuals with lower metacognitive sensitivity show a more robust tendency for stronger moral conviction to be associated with decreased social conformity (Yoder & Decety, 2022). That study used electrophysiological measures and found that metacognitive sensitivity moderates the effects of moral conviction on mid-frontal negativity (MFN), a signal thought to originate in the ACC, when participants evaluate how much they support violent political protesters. Cognitive neuroscience research has found that the spatiotemporal organization of the salience network is predictive of cognitive flexibility (Chen et al., 2016) and the IPFC plays an important role in metacognition (Fleming & Dolan, 2012; Lapate et al., 2020). These regions overlap with some of the regions that we expect to be associated with moral conviction. Thus, it was predicted that individual differences in metacognitive ability may modulate the effects of moral conviction on neural activity during social decision-making.

The current study was designed to determine: 1) the extent to which moral conviction and associated neural activity influence decision time and decision consistency; 2) where and how brain responses are modulated by one's support for and moral conviction about sociopolitical protests during decisions choosing which one of two protests to support; and 3) how individual differences in metacognitive sensitivity influence these neural responses. Based on previous research on moral cognition and value-based decision-making, blood oxygen level-dependent (BOLD) responses in the vmPFC and VS were expected to encode the mean support rating of the two issues presented in each decision, signaling the overall subjective value of the decision. Moreover, the neural activity in the IPFC, amygdala, ACC, and aINS was anticipated to be associated with the moral conviction level of a decision. Metacognitive sensitivity, as an

individual-level trait, was expected to moderate the impact of moral conviction on hemodynamic responses in these regions.

Methods

Participants

Eighty adult US citizens (45 females, 34 males, 1 nonbinary participant; Age range 18 – 48 years; $M_{age} = 23.84$, $SD_{age} = 5.98$) from the Chicago metropolitan area were compensated \$10 to complete an online survey to collect their demographic information and assess their views on current sociopolitical issues. Only those who had 15 or more unique combinations of support and moral conviction scores on the sociopolitical issues were invited to participate in the fMRI study. Forty-nine healthy adult participants completed the fMRI study and were paid an additional \$40 compensation. All participants provided informed written consent, and all procedures were approved by the Institutional Review Board at the University of Chicago. No part of the study procedures or analyses was pre-registered prior to the research being conducted.

Three participants were excluded from both behavioral and fMRI analyses involving the in-scanner task due to a below-chance (less than 50%) level of consistency. A choice was considered consistent when a participant chose the protesters who supported the issue to which they had given a higher support score in the initial survey, and was inconsistent if the opposite was true. Framewise displacement (FD) calculated using MRIQC (Esteban et al., 2017) was used to identify runs and participants with excessive motion. Functional MRI runs for which greater than 5% of volumes had FD above 0.9 mm were excluded from the MRI analyses, as were runs in which the mean FD was at least two standard deviations worse than the mean from all runs. If three or more runs from the same participant met these criteria, all runs from the participant were excluded. Accordingly, one participant and 5 individual runs from 4 participants were excluded

from the MRI analyses due to excessive movement. While we did allow participants who were taking a single SSRI antidepressant to participate in the study, one participant was erroneously recruited despite reporting that they were taking Adderall; this participant was removed from behavioral and fMRI analyses involving the in-scanner task. Metacognitive sensitivity data from 6 participants were lost due to technical issues. Thus, the final behavioral sample for the decision-making task had 45 participants, with fMRI data included for 44 of them (27 females, 16 males, 1 nonbinary participant; Age range 18 – 48 years; $M_{age} = 22.27$, $SD_{age} = 5.06$, 38 of whom provided data on the metacognitive sensitivity measure).

Procedures

The initial survey was completed online at least 1 week prior to the fMRI study. Participants answered 4 questions about a series of sociopolitical issues (Figure 2.1). For each issue, participants indicated their degrees of familiarity (from not at all familiar to very familiar) on a 5-point scale and support (from strongly oppose to strongly support) on a 7-point scale. Moral conviction for each issue was indexed by the average score of two questions using a 5-point scale: “To what extent is your position on ____ a reflection of your core moral beliefs and convictions?” and “To what extent is your position on ____ connected to your beliefs about fundamental right and wrong?” (Skitka & Morgan, 2014). Demographics including age, gender, household income, level of education and religiosity were obtained. Participants provided their political engagement, party registration, party alignment, and political orientation regarding social and economic issues. Religiosity and justice sensitivity were measured by the Duke University Religion Index (DUREL) (Koenig & Büssing, 2010) and the Justice Sensitivity Short Scales (Baumert et al., 2014) respectively.

Prior to functional MRI scanning, participants' metacognitive sensitivity was assessed with a perceptual confidence task designed by Fleming & Lau (2014). The task was programmed using PsychToolbox and run within MATLAB. In each trial, participants saw two circles which contained different numbers of dots and determined which circle had more dots. After each decision, participants rated their confidence level regarding their choice using a five-point scale. The task was self-paced. After 4 example stimuli, the difficulty of the task was customized for

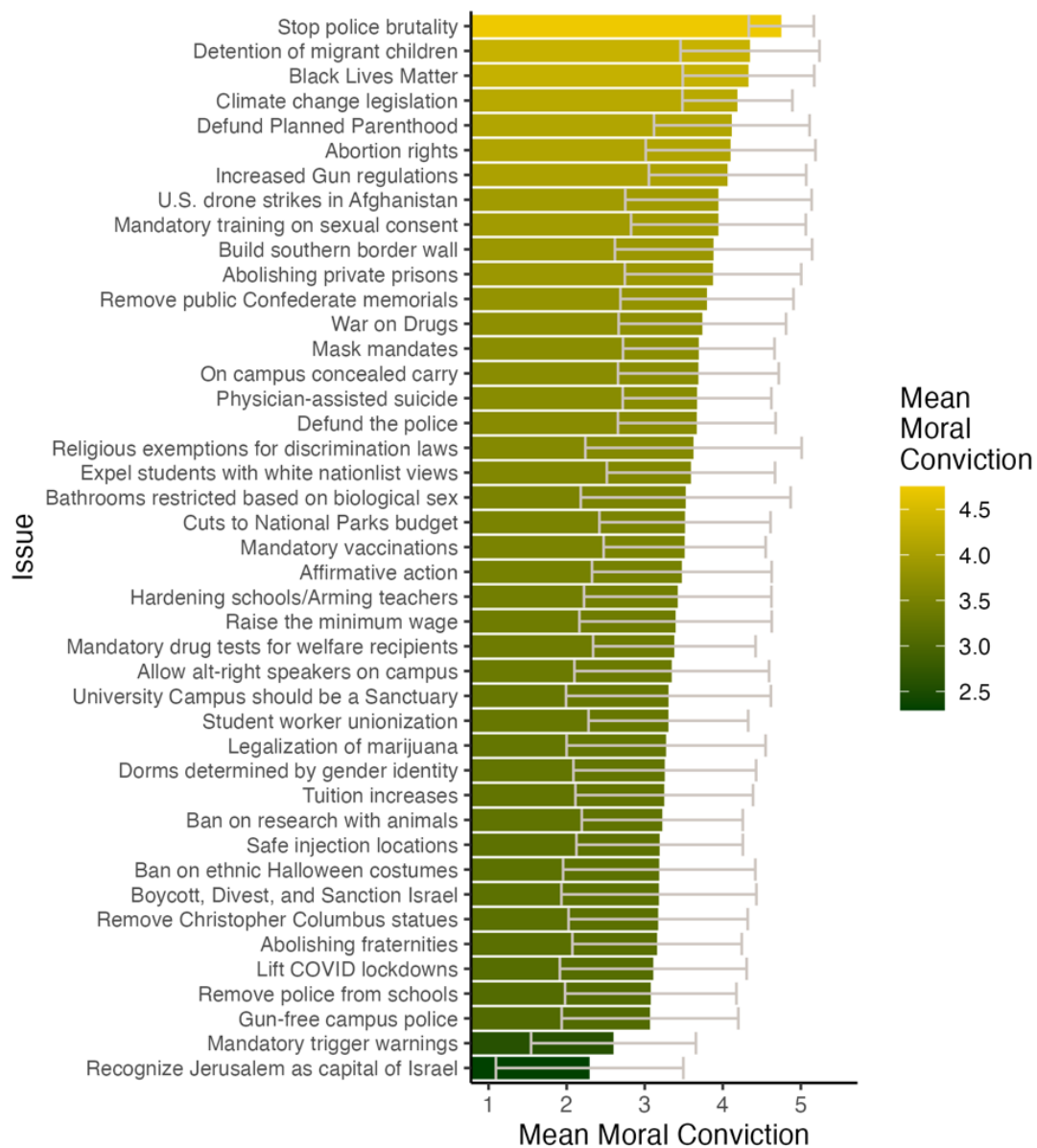


Figure 2.1: Mean moral conviction ratings for each sociopolitical issue from the initial online survey. Error bars represent one standard deviation from the mean.

each participant during an initial calibration phase in which, instead of indicating confidence, each participant received feedback (correct or incorrect) on their judgments. The number of trials in the calibration phase is dependent on each participant. The first calibration trial started with one circle containing 50 dots and the other containing at least 1 and at most 100 dots. The algorithm then flexibly increased difficulty (after 2 correct judgments) or decreased difficulty (after 1 incorrect judgment) by manipulating the number of dots presented in the two circles in a trial. The calibration ends when the difficulty level has reversed 8 times. Supposedly, at this point, the accuracy level reaches approximately 71% for the specific participant doing the task (Levitt, 1971). Data corresponding to the exact numbers of trials in the calibration phase for each participant were not available. The minimum number in theory, based on the algorithm used for calibration, is 12 trials, but often the number of trials was substantially larger than this. After calibration, participants completed 10 practice trials that included confidence judgments, and then two blocks of 25 trials for the main task. Metacognitive sensitivity was computed with a hierarchical Bayesian framework (Fleming, 2017; Maniscalco & Lau, 2012). Each participant's meta-d' was estimated based on their responses, modeled using Markov Chain Monte Carlo (MCMC) as implemented in JAGS (version 3.4.0) within MATLAB.

During fMRI scanning, participants completed a decision-making task (Figure 2.2). In each trial, two photographs showing protesters for or against various sociopolitical issues were presented and participants decided which of the two protest groups they supported more. All issues had been rated previously in the online survey, and unfamiliar issues (familiarity rated not at all familiar) were excluded.

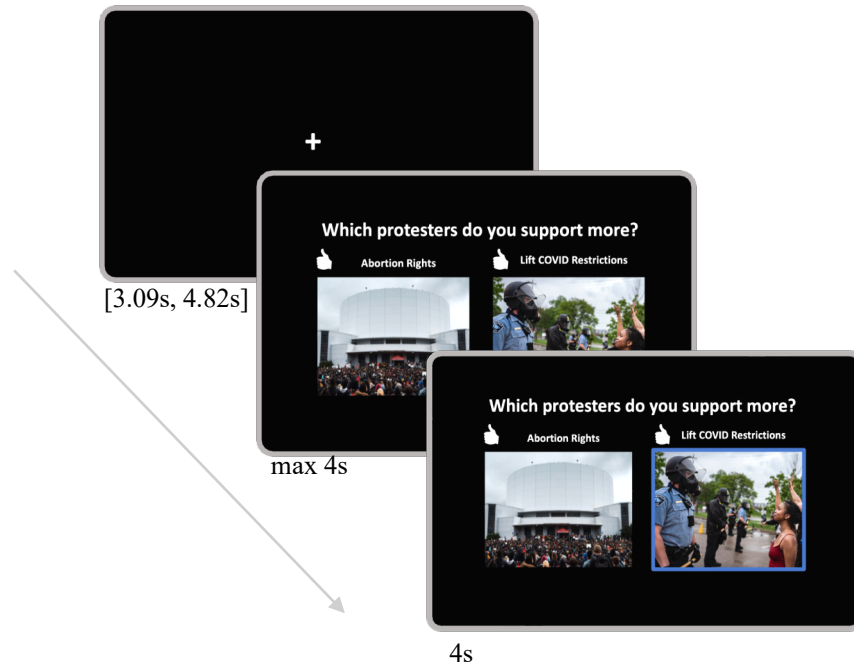


Figure 2.2: Example of stimulus sequence in a trial during fMRI scanning.

A total of 120 trials (5 runs * 24 trials per run) were included. The thumbs-up or thumbs-down icons next to the photos indicated whether the protesters supported or opposed the issue. In other words, a thumbs-down icon reverses the direction of an issue. Both support (thumbs-up) and opposition (thumbs-down) trials were included to achieve a larger range of overall support levels across trials and to reduce the potential collinearity between support and moral conviction ratings. Support or opposition was always consistent for the two issues presented in a single trial and was randomized evenly between trials. Issues shown for each participant were tailored according to their prior ratings so that the support ratings of the two issues within a trial always

differed. If there were more than 120 possible pairs of issues, a randomly selected subset of pairs was used, while if fewer than 120 distinct pairs were possible, some trials were repeated. All familiar issues were presented at least once during the decision-making task. The total number of familiar issues ranged from 24 to 40, with a mean of 36 and an SD of 3.75. Trials were also configured so that the issue with higher previously-rated support was evenly distributed between the right and left sides of the screen. The first trial of each run started with a jittered fixation ranging from 3.09s to 3.91s, while other trials started with a jittered fixation ranging from 3.18s to 4.82s. Participants had 4 seconds to respond before the program would automatically proceed to the next trial. All stimuli were presented in E-prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA, USA).

Behavioral analysis

To examine whether participants' demographic characteristics and dispositional traits affect their general tendency to moralize the sociopolitical issues presented, a mixed-effects linear regression model was fitted with moral conviction as the outcome variable. Intensity of support (calculated as $|\text{support} - 3|$ to capture deviation from the midpoint of the support scale), familiarity with the issue, religiosity, justice sensitivity, gender, age, education, income, party alignment, and political engagement were all included as fixed effects. For all behavioral analyses, gender was coded with two levels: 0 for non-female and 1 for female. Age was standardized. Education was converted to years of education: 12 for *high school diploma or GED or associates or technical degree*; 14 for *some college, but no degree*; 16 for *bachelor's degree*, and 18 for *graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS, etc.)*. Income was coded as a continuous variable: 1 for *less than \$25,000*; 2 for *\$25,000-\$49,000*; 3 for *\$50,000-\$74,999*; 4 for *\$75,000-\$99,999*; 5 for *\$100,000-\$149,999*; 6 for *\$150,000 or more*, and participants who

responded *prefer not to say* were treated as NA. Party alignment was treated as continuous from 0 (*Strong Democrat*) to 6 (*Strong Republican*). Political engagement ranged from 0 (*never*) to 4 (*always*), representing the extent to which participants followed politics and public affairs. Participants and issues were modeled with random intercepts, respectively. Issues with which participants were not at all familiar were excluded from the analysis. While this analysis was conducted with data from all 80 participants who completed the initial online survey, other behavioral analyses relied on data collected during the scan session, so data were available for the 45 participants who took part in the scanning part of the study and met the inclusion criteria for behavioral analyses.

A second behavioral analysis was a mixed effects logistic regression used to test whether the relative protest support rating for the issue presented on the left-hand side of the screen (as compared to the one presented on the right-hand side) predicted a corresponding in-scanner choice of that issue. If an issue was presented in a thumbs-down form, support ratings were multiplied by -1. The relative support was calculated as support rating (left) – support rating (right). The mixed effects model included relative support as the primary fixed effect, with gender, age, education, income, party alignment, and political engagement entered as fixed-effect covariates of no interest. Participants were modeled with random intercepts.

Response time was analyzed as well with a mixed-effects linear regression. The moral conviction level of a trial was operationalized as the higher moral conviction rating of the two issues within a trial and denoted as the maximum moral conviction. In the model, fixed effects included maximum moral conviction, familiarity with the issue having maximum moral conviction, support difference (chosen - unchosen), the protesters' position (thumbs-up vs. thumbs-down), and the interaction between maximum moral conviction and support difference.

As above, gender, age, education, income, party alignment, and political engagement were included as covariates of no interest, and participants were modeled with random intercepts. Only consistent choices were included in the analysis (3961 out of 4823 trials that had responses) to better capture the cognitive processes underlying decisions that reflect participants' stated preferences, as inconsistent trials may involve different cognitive processes, such as increased uncertainty, attentional lapses, or decision error. To confirm the results, an additional mixed-effects linear regression model with the same specifications was fitted to the data from all trials that had responses.

Functional MRI acquisitions and analyses

MRI scanning was conducted with a 3.0 T Philips Achieva MRI scanner (Philips Medical Systems, Best, The Netherlands) equipped with a 32-channel SENSE head coil at the University of Chicago MRI Research Center. First, high-resolution T1-weighted anatomical scans were acquired using a 3D MP-RAGE sequence (TR = 8 ms; TE = 3.5 ms; voxel size = $0.85 \times 0.85 \times 0.85 \text{ mm}^3$; matrix = 284×260). Then, functional images were collected in ascending order and transverse slices using a single-shot EPI sequence with the following parameters: voxel size = $3.0 \times 3.1 \times 3.0 \text{ mm}^3$, flip angle = 80° , matrix = 64×62 , TR = 2000 ms, TE = 28 ms, field-of-view = $192 \times 192 \text{ mm}^2$, slice gap = 0 mm). Each of the 5 runs acquired 135 volumes and lasted 4 mins and 42 s. Participant attention to the task was monitored throughout the scan using an EyeLink 1000 Plus Eye Tracker (SR Research, Ontario, Canada), but eye-tracking data are not reported here.

Preprocessing

Data were preprocessed using fmriprep v22.1.1 (Esteban et al., 2020). The following 4 paragraphs are excerpted and adapted from the documentation distributed with fmriprep. The T1-

weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (part of FSL 6.0.5.1; Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 7.2.0; Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical GM of Mindboggle (Klein et al., 2017). Volume-based spatial normalization to standard space was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was used for spatial normalization: FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model (TemplateFlow ID: MNI152Nlin6Asym; Evans et al., 2012). A B0-nonuniformity map (or fieldmap) was estimated based on two echo-planar imaging (EPI) references with topup (FSL 6.0.5.1; Andersson et al., 2003).

For each BOLD run, the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1; Jenkinson et al., 2002). The estimated fieldmap was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field

coefficients were mapped on to the reference EPI using the transform. BOLD runs were slice-time corrected to 0.975s (0.5 of slice acquisition range 0s-1.95s) using 3dTshift from AFNI (Cox & Hyde, 1997). The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom.

A set of physiological regressors was extracted to allow for component-based noise correction (`aCompCor`) (Behzadi et al., 2007; Muschelli et al., 2014). Principal components were estimated after high-pass filtering of the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off). Probabilistic masks for CSF and WM were generated in anatomical space, and components were calculated separately within each mask.

The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152Nlin6Asym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of `fMRIPrep`. All resamplings can be performed with a single interpolation step by composing all pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Univariate analyses

Data preprocessed in `fMRIPrep` were then analyzed using FSL. Smoothing was applied using a 5 mm full width half maximum (FWHM) Gaussian kernel. At the first level, a general linear model (GLM) containing 5-6 primary regressors and the temporal derivative of each was applied to data

from each scan run. The following primary regressors were modeled: 1) all trials with a response, 2) support mean, 3) support difference, 4) maximum moral conviction, 5) RT, and 6) trials with no recorded response (if present on a given run). Support mean and maximum moral conviction were chosen to represent the overall support level and moral conviction level, respectively, in each trial. The reasoning for using the mean of support ratings for the two causes is to reflect the overall subjective value of a trial. This design allows us to examine whether the support rating a participant assigns to the cause represented by a group of protesters is akin to assigning a valuation to that group of protesters. In other words, if support is akin to value, considering two sets of protesters whom the participant, on average, agrees with would be processed as more rewarding than considering two sets of protesters whose cause is less aligned with the participants' views. The parametric regressor for mean support quantifies this relationship. Cognitive neuroscience studies have found that brain activity in the valuation system, especially the vmPFC, tracks the sum/mean of the subjective value of multiple options in a decision (e.g., Hunt et al., 2012; Levy & Glimcher, 2011). Moral conviction is expected to show a somewhat different pattern: as soon as the decision-maker recognizes one highly morally convicted issue, heightened neural activity in brain regions that play a role in the detection of morally relevant information will be elicited, regardless of the moral conviction level of the other issue. The length of the boxcar was set to 4 seconds for all trials, and each trial was convolved with a double-gamma hemodynamic response function (HRF). Confound regressors representing the following variables, computed in fMRIPrep, were also included in the GLM: 6 motion parameters, 5 aCompCor parameters from WM, 5 aCompCor parameters from CSF (or fewer components if sufficient to account for 50% of variance in the CSF tissue compartment), and 3 cosine basis functions for high-pass filtering. This analysis was intended to identify the brain

areas that track the average support ratings and the maximum moral conviction of the two issues shown.

Regressors computed at the first level were then combined across all runs in a fixed-effects analysis in FEAT, and contrasts were computed, yielding a single regressor for each contrast in each individual. Group-level results were computed in FEAT by combining data from each individual in a FLAME 1 and 2 mixed-effects analysis with automatic outlier detection. A voxel threshold of $z > 3.1$ was applied together with a cluster threshold of $p < 0.05$ implemented using Gaussian random field theory in FEAT.

Region of Interest (ROI) analyses

To examine whether regions of the valuation system are keeping track of the degrees of support and/or moral conviction, ROI analyses were implemented. The ROIs were defined as the vmPFC and VS ROIs identified in a meta-analysis of domain-general reward signal (Bartra et al., 2013). For each participant, activity was calculated by averaging the parameter estimate (COPE) of the parametric effect of support mean and maximum moral conviction from the second level FEAT, across all voxels within the two ROIs.

Functional connectivity analyses

A generalized psychophysiological interaction (gPPI) analysis (McLaren et al., 2012) was conducted to compare functional coupling between specific brain regions for trials varying in maximum moral conviction. Regressors for the gPPI analysis were constructed using AFNI. The average time series in the seed region was extracted from the data after fMRIPrep preprocessing and smoothing in FSL (using a 5 mm FWHM Gaussian kernel). This time series was then detrended and up-sampled by a factor of 20. A gamma function HRF was then deconvolved from the time series using the AFNI 3dTfitter command. This approach used a combination of penalty

functions based on the raw time series and its first and second derivatives, a penalty weight of -2, and Lasso regularization with $\lambda = -6$, following the suggested parameters from the 3dTfitter help file. To create the physiological regressor, the deconvolved time series of the ROI was re-convolved with a gamma HRF and down-sampled back to the TR length. The psychological regressor for gPPI (maximum moral conviction) and 4-5 additional regressors (all trials with a response, support mean, support difference, RT, and missed trials, if present) were computed by up-sampling the raw regressors, convolving them with a gamma HRF, and down-sampling back to the TR length. The gPPI regressors were created by multiplying the up-sampled raw psychological regressor for maximum moral conviction with the deconvolved time series of the seed region, then convolving with a gamma HRF, and finally down-sampling back to the TR length.

For the first-level analysis in FSL, data that were preprocessed in fMRIPrep and smoothed by FSL (using a 5 mm FWHM Gaussian kernel) served as input. No additional preprocessing was done. The primary psychological regressor, physiological regressor, gPPI regressor and the 4-5 additional control psychological regressors listed above were modeled to fit a general linear model. The temporal derivatives for all psychological regressors were calculated and included in the model and temporal filtering was also applied to these regressors. Additional confound regressors computed in fMRIPrep (same as specified in the univariate analysis) were included. A second-level fixed effects analysis combined data from 5 runs, and contrasts of interest were calculated for each participant. Finally, outputs from the second-level analysis were entered as inputs for the group level analysis, which was a FLAME1 and FLAME2 mixed effects model using automatic outlier de-weighting. Cluster thresholding was used with a $z > 3.1$ voxel threshold and a $p < 0.05$ cluster threshold based on Gaussian random field theory.

Univariate effects and behavioral analyses

The univariate analyses mentioned above identified the brain network that was responsive to the moral conviction level on sociopolitical issues. In analyses relating behaviors to brain activity in this network, for each participant, the parametric effects of maximum moral conviction were extracted from each cluster found in the whole-brain parametric modulation analyses. These effects were then averaged (after multiplying activity estimates by -1 in clusters showing negative effects) to form a single measure of moral conviction-related brain activity.

Specifically, a mixed-effects logistic linear regression was fitted to test whether participants who showed a stronger effect of maximum moral conviction in the univariate fMRI analysis also showed more consistency between the initial support rating and in-scanner decisions. Relative support, brain activity related to moral conviction, and the interaction of these two variables were included as the primary fixed effects. Another mixed effects regression was used to examine whether higher maximum moral conviction led to a greater reduction in decision time in those individuals who showed a stronger parametric effect of moral conviction in the univariate fMRI analysis. In this model, support difference (chosen - unchosen), the protesters' position (thumbs-up vs. thumbs-down), familiarity with the issue having maximum moral conviction, maximum moral conviction for the trial, moral conviction-related brain activity, and the interaction between the latter two variables were included as fixed effects. Only trials where the response in the scanner was consistent with earlier support ratings were included for the response time analysis. In both models, participants' gender, age, education, income, party alignment, and political engagement were entered as fixed-effect covariates of no interest. Participants were modeled with random intercepts.

Univariate effects and metacognition analyses

A series of Pearson correlation analyses was conducted to examine the relationships between metacognitive sensitivity and brain activity related to moral conviction and support levels for issues presented in the sociopolitical decisions. Brain activity related to maximum moral conviction or support mean was quantified in the same way described in the previous session, such that the parametric effects of maximum conviction and support mean on fMRI BOLD signal, averaged across all significant clusters, were used as measures of moral conviction and support-related brain activity, respectively. Additional correlation analyses were implemented to test the relationships between metacognitive bias (general tendency to have higher confidence) and the brain activity associated with moral conviction. The Benjamini-Hochberg false discovery rate method was employed in cases where control for multiple comparisons was necessary (Benjamini & Hochberg, 1995).

Results

Behavioral

Results from the mixed-effects linear regression examining effects of demographic characteristics and dispositional traits on moral conviction rating indicated a significant positive association between religiosity and moral conviction [$B = 0.02$, 95% CI (0.003, 0.044), $p = 0.04$] (Figure 2.3). Additionally, regardless of whether the participant supported or opposed the issue, when their position about the issue in question was more extreme, they were more likely to have a higher level of moral conviction, while holding the other variables constant [$B = 0.56$, 95% CI

(0.52, 0.61), $p < 0.001$]. Familiarity positively contributed to moral conviction as well [$B = 0.21$, 95% CI (0.17, 0.26), $p < 0.001$].

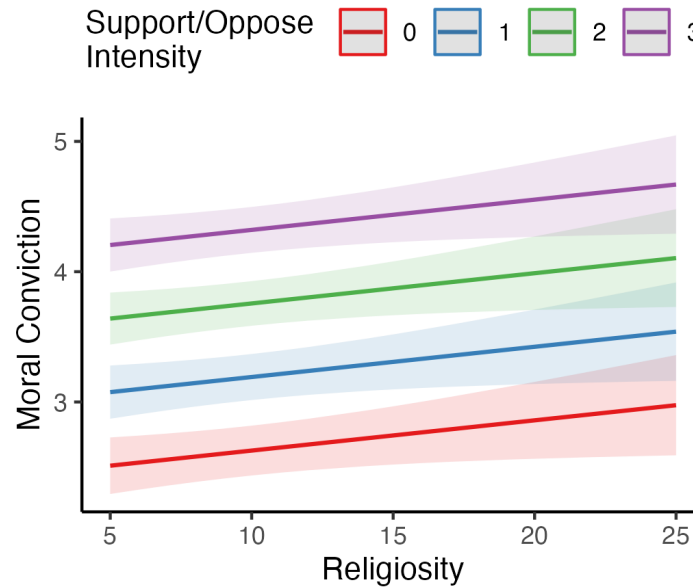


Figure 2.3: Effect of religiosity on moral conviction. The shades illustrate 95% confidence intervals.

Results from the logistic model examining the relationship between relative support and in-scanner choice showed that higher relative support for the issue on the left side of the screen led to a higher chance of choosing to support the protesters for that issue ($p < 0.001$, $OR = 1.82$, 95% $CI = [1.76, 1.88]$, see Figure 2.4). Furthermore, while support ratings from the pre-scan survey were not always perfectly consistent with choices made in the scanner, the pattern of results in Figure 2.4 shows that inconsistent choices were more likely to occur on trials where the support rating difference between the two issues was smaller rather than larger, supporting our assumption that in-scanner choices broadly align with earlier ratings. In a separate mixed effects logistic regression, the main effects of relative support, maximum moral conviction, and their interaction were included as fixed effects, with the same covariates and the random effect. The

interaction between relative support and maximum moral conviction was not significant ($p = 0.46$, $OR = 1.02$, 95% $CI = [0.98, 1.06]$).

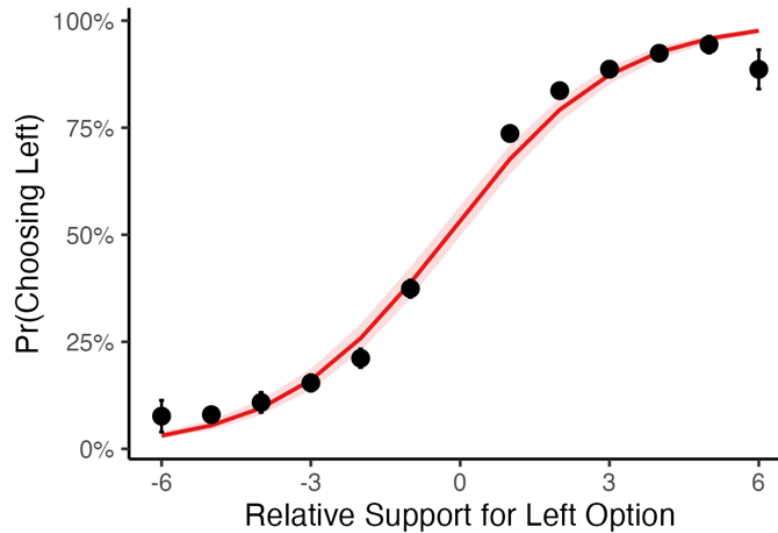


Figure 2.4: Relationship between the probability of choosing to support protesters on the left and the relative support rating of the issue on the left. The shade illustrates the 95% confidence interval.

The mixed effects linear regression analysis demonstrated significantly shorter response times with both higher maximum moral conviction [$B = -69.33$, 95% $CI (-119.36, -19.24)$, $p = 0.007$] and higher support difference [$B = -98.32$, 95% $CI (-164.48, -30.51)$, $p = 0.004$] (Figure 2.5; see Supplementary Reanalysis for an extended model including maximum support).

Participants also made decisions faster when the protesters supported the issue in question (thumbs-up) compared to when they opposed the issue (thumbs-down) [$B = -350.11$, 95% $CI (-389.09, -310.99)$, $p < 0.001$]. Rated familiarity with the more morally convicted issue was associated with quicker decisions as well [$B = -26.98$, 95% $CI (-50.02, -4.65)$, $p = 0.02$].

Numerically, the interaction between support difference and maximum moral conviction trended in the predicted direction, as the association between stronger maximum moral conviction and faster response time was stronger when the support difference was smaller, but this interaction

was not significant [$B = 10.77$, 95% $CI (-4.82, 26.11)$, $p = 0.17$]. In the model where all trials were included, effects of maximum moral conviction [$B = -36.75$, 95% $CI (-64.60, -9.55)$, $p = 0.009$], support difference [$B = -40.55$, 95% $CI (-77.19, -3.91)$, $p = 0.03$], protesters' position [$B = -327.30$, 95% $CI (-362.74, -291.63)$, $p < 0.001$], and familiarity [$B = -34.27$, 95% $CI (-55.10, -13.89)$, $p = 0.001$] were significant and in the same direction as the effects shown in the model with only consistent choices included.

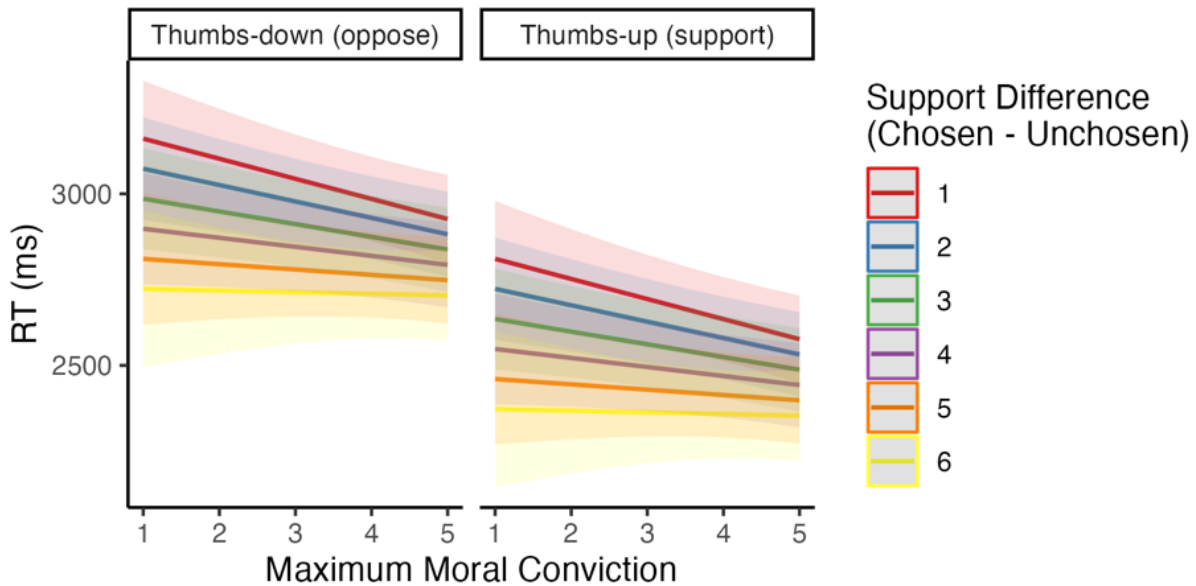


Figure 2.5: Effects of maximum moral conviction, support difference and protesters' position on response time. The shades illustrate 95% confidence intervals.

Neural activity related to support for sociopolitical issues during decision-making

Elevated hemodynamic responses to decisions that had higher mean support ratings were found in the left occipital cortex, vmPFC and left amygdala (Table 2.1, Figure 2.6). The reverse contrast, which identified increased responses to lower mean support ratings, showed significant effects in the dorsal medial prefrontal cortex (dmPFC) and left inferior frontal gyrus (IFG) (Table 2.1, Figure 2.6). All fMRI results in this paper were visualized with the BrainNet Viewer (Xia et al., 2013). An ROI analysis averaging across VS and vmPFC, using regions defined by a

meta-analysis of domain-general reward (Bartra et al., 2013), also showed a significant parametric effect of mean support ($t(43) = 3.00, p = 0.004$). The analyses using VS and vmPFC as separate ROIs indicated a significant parametric effect of mean support in the vmPFC ($t(43) = 3.35, p = 0.002, p_{adjusted} = 0.003$) and not in the VS ($t(43) = 1.32, p = 0.19, p_{adjusted} = 0.19$), controlling for the false discovery rate using the Benjamini–Hochberg method.

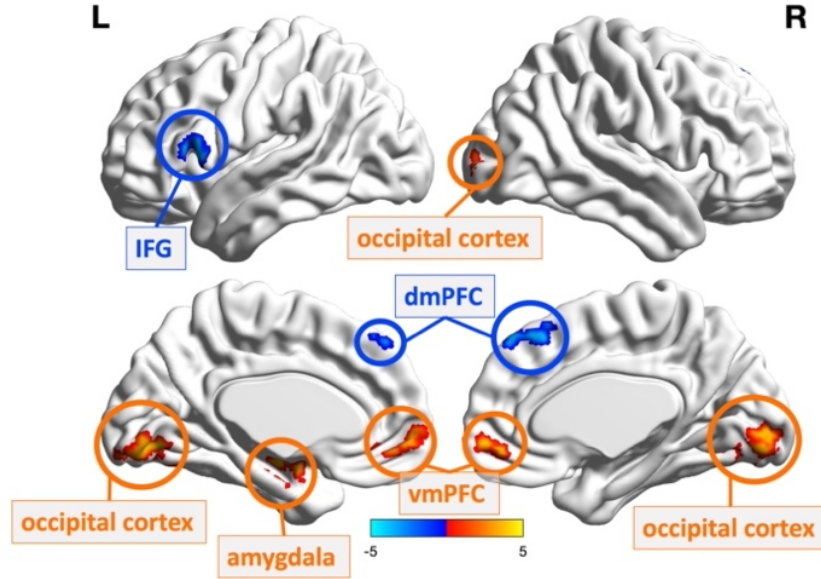


Figure 2.6: Parametric effects of support of sociopolitical issues during decision-making.

Table 2.1: Brain regions showing a significant parametric effect of mean support ratings across the two issues in a given trial.

Brain Region	Cluster Size (Voxels)	Peak MNI Coordinates			Peak z statistic
		x	y	z	
Greater Response					
Left Occipital	760	-10	-74	-6	4.85
vmPFC	255	-2	50	-10	4.1
Left amygdala	120	-20	-4	-24	4.56
Hippocampus		-28	-10	-28	3.91
Weaker Response					
dmPFC	112	0	44	48	4.27
Left IFG	88	-54	20	8	4.27

Neural activity related to the moral conviction level of sociopolitical issues during decision-making

This parametric modulation analysis showed brain regions within which BOLD signal was greater or weaker based on the maximum moral conviction in a given trial. Left inferior frontal cortex, pre-supplementary motor area (SMA), IPFC, and bilateral aINS showed increased hemodynamic responses to trials with higher maximum moral conviction, while the left precuneus displayed decreased activity on trials with higher maximum moral conviction (Table 2.2, Figure 2.7). An ROI analysis examining the role of VS and vmPFC in moral conviction (parametric effect of maximum moral conviction, averaged across these regions) was not significant ($t(43) = -0.79, p = 0.43$). ROI analyses examining the VS and vmPFC separately showed a marginal trend towards a negative effect of maximum moral conviction on activity in the VS ($t(43) = -2.06, p = 0.046, p_{adjusted} = 0.09$) and no significant impact in the vmPFC ($t(43) = 0.40, p = 0.69, p_{adjusted} = 0.69$).

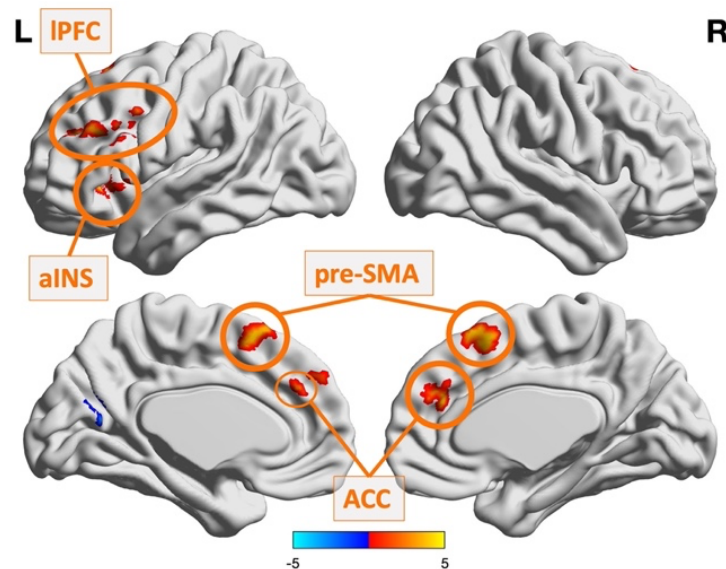


Figure 2.7: Parametric effects of moral conviction of sociopolitical issues during decision-making.

Table 2.2: Brain areas showing significant parametric effects of maximum moral conviction rating on hemodynamic response.

Brain Region	Cluster Size (Voxels)	Peak MNI Coordinates			Peak z statistic
		x	y	z	
Greater Response					
pre-SMA	234	0	16	58	4.39
Left inferior frontal cortex	230	-50	10	32	4.37
Left aINS	130	-42	20	-4	4.44
ACC	87	0	36	30	3.86
Left IPFC	70	-34	46	24	4.41
Weaker Response					
Left precuneus	79	-10	-64	22	4.28

Moral conviction and functional connectivity using IPFC as the seed region

The cluster in IPFC that showed a significant response to maximum moral conviction in the univariate analysis was selected as the seed region for a gPPI analysis. This choice was based on the LFPFC's pivotal role in complex cognitive functioning and value-based decision-making.

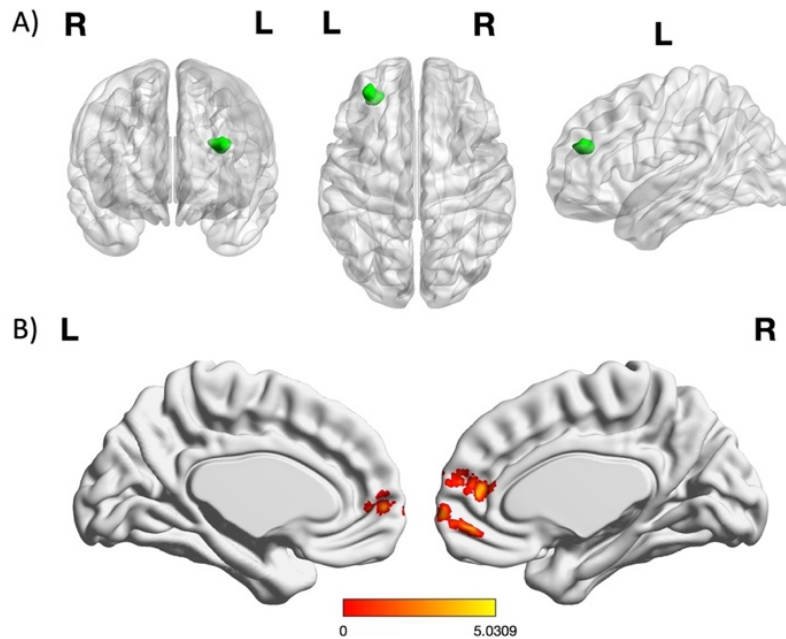


Figure 2.8: A) IPFC seed region and B) gPPI connectivity higher maximum moral conviction.

Prior studies have suggested that IPFC interacts with brain regions including the vmPFC, ACC, aINS, and the posterior cingulate cortex (PCC) to achieve various goals, including those that are morally relevant (Carlson & Crockett, 2018; Dixon & Christoff, 2014; Duverne & Koechlin, 2017). Thus, psychophysiological interaction analysis was conducted to examine whether functional connectivity with IPFC—specifically with the cluster responsive to moral conviction—varied depending on the moral conviction level of a sociopolitical decision. The analysis demonstrated that, during decisions with higher maximum moral conviction, there is significantly stronger functional connectivity between the IPFC seed cluster and the vmPFC and mPFC (Figure 2.8, Table 2.3).

Table 2.3: Brain regions that showed significant increases in functional connectivity with IPFC during decisions with higher maximum moral conviction.

Brain Region	Cluster Size (Voxels)	Peak MNI Coordinates			Peak z statistic
		x	y	z	
vmPFC	224	0	52	-6	4.62
mPFC	141	4	52	14	3.97

Neural activity related to moral conviction, choice consistency, and decision time

The logistic regression model used to examine how relative support and the parametric effect of moral conviction on brain activity together predict in-scanner choices showed that the relative level of support for an issue in the pre-scan survey was associated with a higher chance of choosing to support the protesters corresponding to that issue ($p < 0.001$, $OR = 1.70$, 95% $CI = [1.63, 1.78]$). A significant interaction was found between this effect and the participant's parametric effect of moral conviction on brain activity during these choices ($p < 0.001$, $OR = 1.02$, 95% $CI = [1.008, 1.024]$, see Figure 2.9A). Specifically, individuals whose neural

responses were more strongly modulated by moral conviction showed greater alignment between the pre-scan support ratings and in-scanner decisions.

The mixed effects linear regression model predicting decision time showed that individuals with a greater parametric effect of moral conviction on brain activity had slower response times overall [$B = 34.91$, 95% CI (4.67, 64.30), $p = 0.03$], but this effect was modulated by a significant interaction with trial-level maximum moral conviction [$B = -7.99$, 95% CI (-14.08, -1.70), $p = 0.01$]. Specifically, in participants with stronger parametric modulation of neural responses to moral conviction, higher maximum moral conviction was associated with faster decisions, while this effect was not observed in those with weaker parametric modulation in response to moral conviction (Figure 2.9B). Shorter response times were also associated with a larger difference in support between the chosen and non-chosen choice options [$B = -50.11$, 95% CI (-63.20, -36.39), $p < 0.001$], greater familiarity with the issue having higher moral conviction [$B = -30.41$, 95% CI (-53.67, -8.01), $p = 0.01$], and when protesters are in support (thumbs-up) rather than opposed to (thumbs-down) the issue in question [$B = -345.73$, 95% CI (-385.14, -306.13), $p < 0.001$].

Meta-d' and parametric effects of moral conviction and support of sociopolitical issues

No significant correlation was found between an individual's metacognitive sensitivity (measured by meta-d') and support-related brain activity [$r(36) = 0.19$, $p = 0.24$], but a significant correlation between metacognitive sensitivity and maximum moral conviction was observed (Figure 2.10A) [$r(36) = -0.38$, $p = 0.02$]. To confirm that this observed effect was not being driven by a relationship between overconfidence and dogmatic belief, the mean confidence rating on the perceptual confidence task was calculated for each participant as a measure of metacognitive bias and was correlated with brain activity related to maximum moral conviction.

No significant correlation was found [$r(36) = -0.14, p = 0.39$]. An effect driven by overconfidence would imply a positive correlation, which clearly is not apparent here. Finally, to clarify whether the neural response in specific clusters drove the negative association between meta-d' and the mean parametric effect of moral conviction, Pearson correlation analyses were carried out to test the relationship between metacognitive sensitivity and the parametric effect of moral conviction in each of the 6 clusters. Results showed that meta-d' was negatively associated with the neural activity in the left IPFC, [$r_{IPFC}(36) = -0.44, p = 0.01, p_{adjusted} = 0.03$], and ACC [$r_{ACC}(36) = -0.35, p = 0.03, p_{adjusted} = 0.09$], though the latter effect was not significant after controlling for the false discovery rate. No significant effects (uncorrected $p < 0.05$) were found in any other areas. These results suggest that IPFC, and potentially ACC, are driving the effects across the full network.

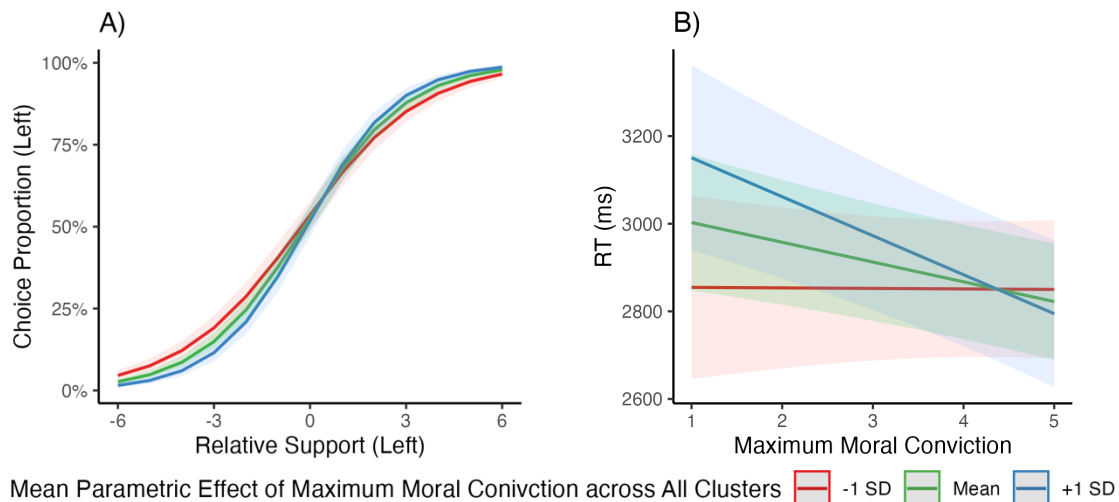


Figure 2.9: A) Relationship between probability of choosing to support protesters on the left-hand side of the screen and the relative support rating of the issue from the pre-scan survey, moderated by the strength of an individual's parametric effect of moral conviction on brain response in regions shown in Figure 2.7. B) Relationship between decision time and maximum moral conviction on each trial, moderated by the individual's parametric effect of moral conviction on brain response in regions shown in Figure 2.7. Shadings indicate 95% confidence intervals.

The role of valuation regions was further examined by correlating the parametric effects of moral conviction and support in the valuation network (by averaging responses from the VS and vmPFC) with metacognitive sensitivity. A significant negative correlation was found between metacognitive sensitivity and the parametric effect of moral conviction in reward-sensitive regions [$r(36) = -0.38, p = 0.02$] (Figure 2.10B). When examining the two reward-sensitive regions separately, a significant negative association between metacognitive sensitivity and the parametric effects of moral conviction in the VS was found [$r(36) = -0.32, p = 0.0496, p_{adjusted} = 0.06$], though this effect was not significant after controlling for false discovery rate. The association between metacognitive sensitivity and the parametric effect of moral conviction in the vmPFC also trended in the negative direction [$r(36) = -0.31, p = 0.06, p_{adjusted} = 0.06$]. No significant correlations were found between metacognitive sensitivity and the parametric effect of mean support in the valuation network (either for VS and vmPFC individually or for their average) (combined: $r(36) = -0.13, p = 0.42$; VS: $r(36) = -0.14, p = 0.41, p_{adjusted} = 0.59$; vmPFC: $r(36) = -0.09, p = 0.59, p_{adjusted} = 0.59$).

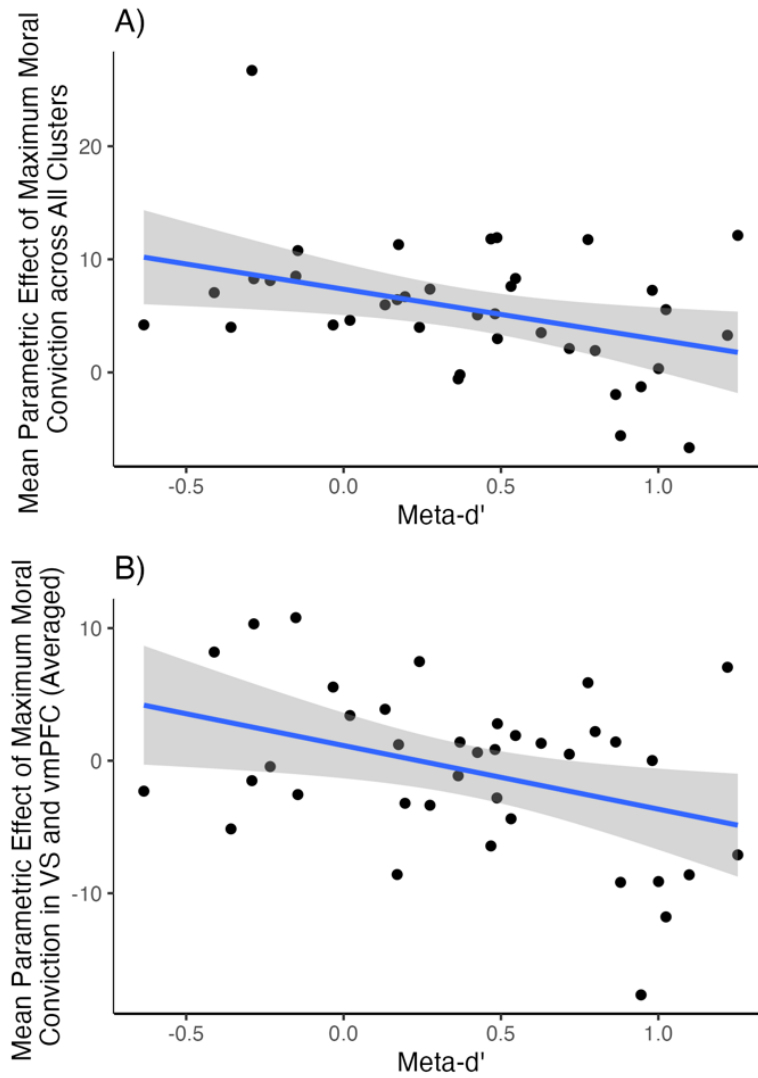


Figure 2.10: Negative correlations between meta-d' and the parametric effects of moral conviction A) across all responsive regions and B) in cortical valuation regions (VS and vmPFC).

Discussion

Moral convictions are perceived as absolute, universal, and definite beliefs or principles. They can operate as moral imperatives that delineate which opinions, actions, and policies are right or wrong, and can motivate collective actions (Decety, 2024). By integrating fMRI data and behavioral measurements, this study provides new evidence about the neural and cognitive underpinnings of moral conviction, including its relationship with metacognitive abilities.

In keeping with previous research in social psychology (e.g., Van Bavel et al., 2012), the behavioral data suggest that moral conviction serves as an indicator of choice significance, as it leads participants to make faster decisions about highly moralized items. These effects persisted when controlling for the difference between the support levels of the two protest groups and the level of familiarity with the highly moralized issue, demonstrating that moral conviction is more than just attitude strength or familiarity. While previous research has shown that moral evaluations are faster than non-moral ones (Van Bavel et al., 2012; Van Berkum et al., 2009), the current study addresses how different degrees of moral conviction affect decision-making about timely sociopolitical issues that have strong real-life implications and shows that content that evokes high levels of moral conviction is processed faster.

A positive association was found between participants' religiosity and their moral conviction ratings of the issues presented, controlling for participants' support level, justice sensitivity, age, gender, education, income, party alignment, and political engagement. This finding is consistent with past work showing that religion promotes *righteous* morality over *prosocial* morality, and with prior findings that religious morality is primarily deontological and non-consequentialist (Saroglou & Craninx, 2021). While some studies have found evidence for a distinction between moral and religious convictions (e.g., Skitka et al., 2018), others suggest that

they do not differ categorically (Brownlee, 2017; Cousar et al., 2021). Results from the current study support that participants with high religiosity potentially have a greater tendency to think about sociopolitical issues from a righteous point of view, which leads to higher overall moral conviction. Further investigation is needed to clarify the relationships between moral conviction and religious attitudes.

Consistent with our theoretical perspective, whole-brain univariate analyses suggest a twofold neurocognitive process underlying moral conviction. An emotional component is reflected by increased neural activity in the salience network, which includes the insula and ACC, in response to high moral conviction. This component signals the salience of morally convicted items, which may subsequently regulate downstream cognitive functions that underlie moral reasoning and decisions. This result aligns with studies showcasing the importance of the salience network in the detection of morally charged information, as well as with the role played by the ACC and the inferior frontal gyrus in distinguishing moral vs. conventional norms (Eres et al., 2018; Sevinc et al., 2017; White et al., 2017). The results also indicate that the cluster in IPFC is more activated in the context of high levels of moral conviction, possibly supporting the cognitive dimension of moral conviction. Previous research showed the implication of IPFC in various functions, including cognitive control, executive function, planning, social cognition, and moral judgment (Forbes & Grafman, 2010; Miller & Cohen, 2001). When it comes to moral cognition, one study reported that disruption of the right dlPFC by transcranial magnetic stimulation causes participants to act less fairly toward a social partner (Knoch et al., 2009). Another study showed that patients with dlPFC damage are less likely to cooperate in a public goods game (Wills et al., 2018). Together, these studies suggest that the dlPFC exerts self-control and inhibits selfish behaviors. In other contexts, however, increased neural response in

the IPFC has been associated with dishonest and selfish behaviors (FeldmanHall et al., 2012; Greene & Paxton, 2009). Thus, results from the present study add to our knowledge of the role of the IPFC, demonstrating that it may not be limited to inhibiting impulsive behavior, but instead flexibly engages in upholding and enforcing goals (Carlson & Crockett, 2018; Tusche & Hutcherson, 2018). The cognitive dimension of moral conviction reflects the ability to distinguish morally convicted beliefs from beliefs lacking moral conviction and underscores their perceived objectivity, grounded in universal and unalterable facts that transcend personal and social boundaries (Wright et al., 2008). This cognitive aspect shapes moral goals and serves as the key to aligning actions with these aims. This conceptual overlap between achieving goals and the imperative nature of moral conviction suggests that the IPFC is a compelling candidate region for implementing the cognitive dimension of moral conviction.

While behavioral ratings of moral conviction showed no significant effect on the consistency between pre-scan support rating and in-scanner decisions, neural activity associated with moral conviction did moderate the relationship between pre-scan support ratings and these same in-scanner decisions. Specifically, individuals who exhibited a stronger parametric relationship with moral conviction in the brain regions identified across the sample (see Figure 2.7) also showed greater consistency between their pre-scan survey responses and their choices in the scanner. The strength of the same parametric brain response to moral conviction also modulated the relationship between moral conviction and decision time. Higher moral conviction was associated with shorter decision times, particularly in those individuals whose brain responses were more strongly modulated by moral conviction. These results suggest that individual differences in the neural encoding of moral conviction reflect moral significance to a degree that behavioral ratings cannot, and highlight the unique contribution that neural measures

can make towards explaining judgment and decision-making, beyond what is possible with purely behavioral and self-report measures.

A gPPI analysis using as the seed region the IPFC cluster parametrically modulated by maximum moral conviction in the univariate analysis demonstrates stronger functional connectivity with vmPFC and mPFC during decisions with higher moral conviction. In cognitive tasks, the functional coupling between dlPFC and vmPFC has been linked to the successful exertion of self-control in choosing larger-delayed rewards over smaller-immediate ones and in choosing healthier over tastier food items (Hare et al., 2009, 2014). Dorsolateral PFC and vmPFC interactions have also been shown to contribute to adaptive value calculations in different contexts, in the absence of a requirement for self-control (Rudorf & Hare, 2014). Therefore, it is likely that the IPFC-vmPFC connectivity signals an increase in the importance of domain-general goals (e.g., health goals, moral goals) and greater incorporation of these goals into the value-based decision-making process. Taken together with the univariate results that the IPFC tracks moral conviction while vmPFC and amygdala track overall support, the increase in functional coupling between IPFC and vmPFC during higher moral conviction decisions suggests an increased integration of moral considerations and sociopolitical opinions. In line with this interpretation, one previous study examining fairness and costly punishment found increased connectivity between right dlPFC and posterior vmPFC in people who more frequently decided to exert costly punishment (Baumgartner et al., 2011), indicating that such neural connectivity may underlie the enforcement of moral conviction related to fairness.

Metacognitive sensitivity moderated neural responses to moral conviction level during decision-making, particularly in left IPFC and potentially in ACC. Our data provide no evidence of a positive correlation between metacognitive bias and neural responses related to moral

conviction, indicating that the relationship with moral conviction is specific to low metacognitive sensitivity and not overconfidence. In previous literature, the LPFC has been implicated in metacognitive judgments (Fleming & Dolan, 2012; Lapate et al., 2020). The present study is the first to demonstrate that metacognition and moral conviction may rely on overlapping circuitry in LPFC. Although the finding that metacognitive sensitivity moderated neural response to moral conviction in ACC was no longer significant after correction for multiple comparisons, such an effect would be consistent with previous findings. Specifically, reduced metacognitive sensitivity has been associated with stronger medial frontal negativity (MFN) elicited as a function of moral conviction level for highly moralized issues (Yoder & Decety, 2022), and the MFN is thought to originate in the ACC (Gehring & Willoughby, 2002). Together, these findings contribute to a growing body of evidence highlighting the role of the LPFC and ACC in metacognitive abilities and moral decision-making. Future research is needed to investigate the underlying neural mechanisms at a more granular level.

Moral conviction showed no significant main effect on activity in the valuation system (vmPFC and VS). Interestingly, though, the magnitude of this effect across individuals was negatively associated with metacognitive sensitivity. In other words, compared to individuals with higher metacognitive sensitivity, those with lower metacognitive sensitivity exhibited greater activity in the vmPFC and VS during trials with higher maximum moral conviction. Thus, moral conviction may be seen as rewarding specifically for those with poor metacognitive abilities. Meanwhile, results from the whole-brain univariate analyses and ROI analyses demonstrated that greater mean support for social issues was associated with greater activity in vmPFC and amygdala. These are key regions in the valuation system, suggesting that moral issues with which one agrees are among the many domains in which liking is indexed in the

reward circuit (Hare et al., 2008; Mormann et al., 2019). The effects of mean support in the reward system, though, showed no significant correlation with metacognitive sensitivity. Overall, these results provide evidence that the extent to which the valuation system is involved in moral conviction may depend on individuals' metacognitive abilities, while its role in tracking the level of support is more universal. Past behavioral studies have linked lower metacognition to dogmatism, political radicalism, and reduced social conformity (Osorio T. & Reyes M., 2023; Rollwage et al., 2018; Yoder & Decety, 2022). The current study suggests a possible neural mechanism through which metacognition regulates moral conviction's impact.

One limitation of the current study is that multiple cognitive processes may have occurred simultaneously, increasing the difficulty of unambiguously relating brain activity to specific cognitive processes. For instance, to what extent is attitude strength distinct from moral conviction? Does the brain's valuation system simply track attitude strength and how rewarding it is to act on a morally convicted attitude, or does it also distinguish between actions that are morally neutral and those driven by strong moral conviction? Moreover, higher moral conviction towards an issue may lead to differences in key sub-processes of decision-making, such as working memory load, attention allocation, and evidence accumulation, all of which are important for metacognition. Future studies should be designed to clarify the basic cognitive processes that play a role in moral conviction, to better understand the neurocognitive mechanisms by which metacognitive sensitivity influences morally charged sociopolitical decisions. In addition, while emotion is a characteristic of moral beliefs and moral conviction, emotional salience does not necessitate moral conviction. Additional questions remain, to be more thoroughly addressed in future work about this distinction (Avramova & Inbar, 2013; Wright et al., 2008).

Moral values and social norms are crucial for fostering cooperation, which is the cornerstone of human civilization. Moral conviction specifically influences opinions, attitudes, and behaviors, leading to a range of outcomes from positive collective action to dogmatism, intolerance, and violence (Decety, 2024). Past studies have established a link between strong moral conviction and social and political intolerance, which exacerbates the distinction between “us” and “them”, and potentially leads to the approval of sociopolitical violence aligned with one’s own ideological views (Garrett & Bankert, 2020; Pretus et al., 2023; Workman et al., 2020; Yoder & Decety, 2022). This study builds on existing knowledge by delineating the mechanisms through which support, moral conviction, and metacognitive sensitivity guide decisions about sociopolitical protests and provides a basis for future research investigating the psychological roots of political and social action.

Conclusion

The study sheds light on the neural mechanisms underlying how moral conviction guides decisions on sociopolitical issues and interacts with metacognitive abilities. Moral conviction is associated with greater neural activity in core regions of the salience network (i.e., ACC and aINS), as well as in regions associated with executive functioning (IPFC and pre-SMA). These neural mechanisms support the emotional and cognitive dimensions of moral conviction, respectively. We propose that regions in the salience network encode the emotional intensity of morally convicted issues while the IPFC plays an important role in recognizing moral goals and signaling the objectivity of moral conviction. Increased coupling between the IPFC and vmPFC as a function of moral conviction, as revealed by a gPPI functional connectivity analysis, suggests a potential decision mechanism by which moral goals tracked by the IPFC are incorporated into the valuation process to a greater extent when a decision involves sociopolitical

issue(s) with stronger moral conviction. The neural responses associated with moral conviction are also stronger in individuals who score lower on metacognitive sensitivity. Activity in vmPFC and VS associated with moral conviction, although not significant in the aggregate, is stronger in individuals with lower metacognitive sensitivity as well. These findings may provide a mechanistic explanation for the increasingly documented observation in cognitive science that beliefs tend to be more rigid in people with poor metacognitive performance. Activity in reward circuitry tracks the value of the decision options and is associated with the overall level of support for the two decision options. This effect is uncorrelated with metacognitive sensitivity, suggesting that activity in the valuation system tracks support regardless of metacognitive abilities. Overall, the study provides new insight into the cognitive and neural mechanisms of moral conviction and provides a basis for further examination of the interplay between moral conviction and metacognition.

Appendix A

The Role of Support Strength in Response Time

To further clarify whether the faster response times were driven by relative support strength between the two options or by the absolute support strength for the favored option, an additional mixed-effects linear regression model was estimated. In the published analysis, maximum moral conviction and support difference were the major variables of interest in predicting response time. However, support difference alone does not distinguish between trials in which one option is strongly supported and those in which none is strongly supported but differ by the same amount. To test whether responses were quicker when at least one option is strongly supported, the present reanalysis added maximum support, defined as the higher support rating of the two options presented on a given trial, as a fixed effect. Everything else remained the same as the published analyses, and this model was fitted to data where the response in the scanner was consistent with earlier support ratings. To determine whether including the maximum support improved model fit, the original model was compared with this extended model. Because the two models had different fixed effects, they were fit using maximum likelihood estimation (REML = FALSE) and compared using a likelihood-ratio test. The extended model provided a significantly better fit to the data. For parameter estimation, the final model was refit using restricted maximum likelihood (REML = TRUE). Results showed a significant negative association between maximum support and response time [$B = -99.73$, 95% $CI (-115.59, -83.91)$, $p < .001$]. In this model, however, the support difference between the chosen and unchosen option did not show a significant impact on response time [$B = -47.69$, 95% $CI (-113.64, 18.26)$, $p = .156$]. The negative effect of maximum moral conviction on response time remained comparable to the published results [$B = -65.30$, 95% $CI (-114.24, -16.37)$, $p = .009$]. Familiarity of the option that

had maximum moral conviction and whether the protesters supported or opposed the issues also had significant and similar effects as found in the published results (familiarity: $B = -24.65$, 95% $CI (-46.84, -2.47)$, $p = .029$; support/oppose: $B = -329.92$, 95% $CI (-368.22, -291.63)$, $p < .001$). These results suggest that faster responses were more strongly related to the absolute strength of support for the favored option than to the relative support gap between chosen and unchosen options. This pattern is consistent with the interpretation that once people see an issue that they strongly support, the decision can become easy, and they make faster decisions. Notably, in both the published results and the current reanalysis, moral conviction accelerated decisions, converging on the robust conclusion that highly moralized content is processed more quickly, which can potentially reduce deliberation and contribute to greater intolerance toward disagreeing others.

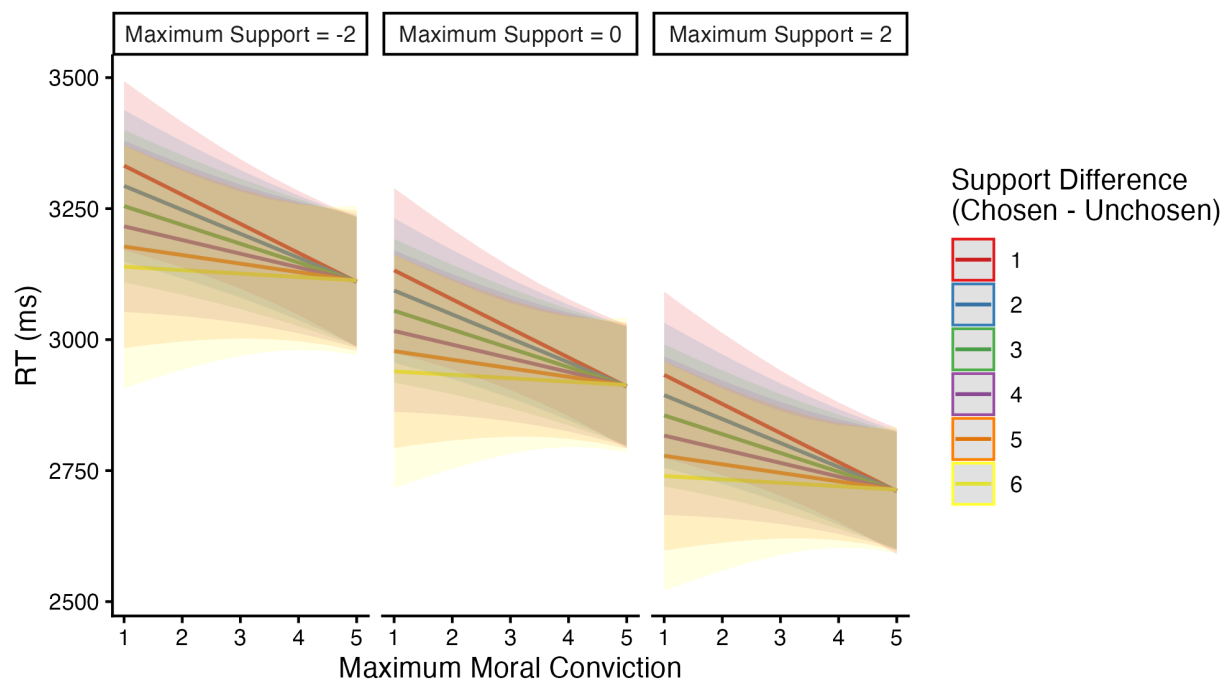


Figure 2.11: Effects of maximum moral conviction, support difference, and maximum support on response time. Note that only maximum moral conviction and maximum support had a significant impact on response time, whereas support difference did not. The shades illustrate 95% confidence intervals.

CHAPTER 3

ATTENTION AND COGNITIVE MECHANISMS IN SOCIOPOLITICAL DECISION-MAKING ABOUT MORALLY CONVICTED ISSUES¹

Accompanied by the prevalence of social media and political polarization, moral conviction has attracted increasing attention in the United States over the past few decades. More concepts, figures, and topics have become moralized, and people are sharing such moralized content more often than neutral content. Yet at the individual level, little is known about how moral conviction shapes attention during real-time decision-making. This chapter uses eye-tracking to investigate whether moral conviction guides visual attention when individuals judge their support for protesters associated with sociopolitical issues that vary in moral conviction. It also tests how moral misalignment, the extent to which a belief-incongruent position is moralized, affects attention and decision-making processes using a Drift Diffusion Model (DDM), while also accounting for related variables such as attitude strength. The results showed that, even after controlling for attitude strength, higher moral conviction was associated with greater dwell proportion. These findings suggest that moral conviction not only influences how individuals interact with others but also how they engage with information.

¹ Chapter 3 was primarily written by the author. The abstract, introduction and discussion were written entirely by the author. Most of the methods and results sections were also written by the author, with some portions about moral misalignment adapted from text originally drafted by Abdulaziz AlHothi, who is a colleague in the 2024-2025 Master of Arts Program in the Social Sciences. Abdulaziz AlHothi contributed to the project through data preprocessing and by conducting the subset of the analyses about moral misalignment included in this chapter. These contributions are incorporated here because of their relevance to the dissertation.

Introduction

Response time analyses provide us insights into how moral conviction, support strength, and familiarity toward the sociopolitical topics slow or speed up decisions. Eye-tracking provides additional information on the underlying processes of choices (Fiedler & Glöckner, 2015).

Previous research has shown that both bottom-up and top-down processes can shift moral decisions. Specifically, when making moral judgments such as agreeing or disagreeing with the statement “Murder is sometimes justifiable”, manipulating the last fixation and the timing of the decision (bottom-up influences) systematically biases people’s choices toward the last-fixated option (Ghaffari & Fiedler, 2018; Pärnamets et al., 2015). Self-determined last-fixated option (top-down influence) is also associated with higher chances of choosing the last-fixated option in a moral decision. These results indicate that attention can both guide and reflect the cognitive processes underlying important moral decisions and suggest that eye-tracking can be a useful tool for studying decisions involving moralized issues.

As discussed in Chapter 2, moral conviction comprises both cognitive and affective dimensions. Cognitively, people believe their morally convicted positions are almost objectively true, and affectively, they associate those positions with intense emotions (Wright et al., 2008). Research has established the importance of emotions in the formation of moral conviction (Wisneski & Skitka, 2017). In particular, both emotionally charged persuasive framing and episodic framing (i.e., concrete and personal framing) can increase moralization by eliciting anger and disgust (Ciuk & Rottman, 2021; Clifford, 2019). Beyond the initial moralization process, emotion intensity remains an important component of established moral conviction. This claim is supported by behavioral, physiological, and neuroscience research. Behaviorally, one study found that US adults whose support or opposition to the Iraq War was high in moral

conviction also had higher emotional reactions to the war, even when their attitude strength was controlled (Skitka & Wisneski, 2011). The same study also found that high moral conviction on physician-assisted suicide (PAS) was associated with greater willingness to engage in activist behaviors, and that this effect was partially mediated by negative emotions among PAS opponents and positive emotions among PAS supporters. Physiologically, moral conviction about political issues, parties, and leaders predicted greater arousal, which corresponds to higher emotional intensity; whereas attitude extremity or importance did not (Garrett, 2019). Previous neuroscience studies have also supported the strong link between moral conviction and emotions. For example, EEG research showed that components indexing early automatic and affective processing (EPN) and saliency and valuation (MFN) were enhanced in response to moralized issues (Yoder & Decety, 2022). Relatedly, results in Chapter 2 converge and demonstrate that stronger moral conviction elicited greater hemodynamic responses in regions in the salience network.

While little research has established the direct relationship between moral conviction and attention captured by oculomotor behaviors, a rich body of research has shown that emotional stimuli automatically capture attention, especially early attention (Calvo & Lang, 2004; Pilarczyk & Kuniecki, 2014; for review see Mulckhuyse, 2018). Calvo and Lang (2004) found that both pleasant and unpleasant pictures attract higher probability of the first fixation and the proportion of viewing time during the first 500ms; Pilarczyk and Kuniecki (2014) further demonstrated that emotional images attract early attention primarily because of their semantic meaning (e.g., emotional relevance), rather than their low-level visual saliency features such as local contrasts of luminance and opposing colors. Eye-tracking studies, as such, used various types of stimuli to capture the valence and intensity of emotions, including images of human

faces expressing different emotions, scenes depicting pleasant and unpleasant events, and animals that are dangerous vs. cute.

Although less evidence is available for the link between the cognitive dimension of moral conviction and attention compared with the well-documented link between emotional stimuli and attention, some studies support the idea that moral content may attract attention for reasons beyond its emotional features. Specifically, in an ERP study where brain activity was measured during a lexical decision task, Gantman and colleagues (2020) found that both arousal and valence contributed to P3 amplitudes, a neural marker associated with attentional engagement and memory (Polich, 2007). Importantly, when each of these factors was modeled simultaneously with word category (moral vs. non-moral), the moral category still accounted for unique variance in brain activity beyond the variance explained by arousal or valence.

Extending beyond mechanistic work in psychology and neuroscience, recent research has increasingly examined moralization at a broader societal level, particularly the moralization of previously non-moral topics and the growing prevalence of moralized content in public discourse and on social media. A recent study analyzing more than 20,000 concepts in the Corpus of Historical American English over the past 150 years and the New York Times annotated corpus from 1987 to 2007, found that various concepts in the diseases and sciences categories have become more morally loaded across time (Ramezani et al., 2026). Relatedly, temporal analyses of social media discourse have found a sharp rise in moralized language from 2013 to 2021 (Puryear et al., 2025). These phenomena have been attributed to many factors. For example, one group of researchers has proposed a MAD model, which states that the accelerated spread of moralized content is supported by *group-identity-based motivation*, greater *attention* captured by moralized content, as well as the *design* of social-media platforms (Brady, Crockett, et al., 2020).

Empirically, in an attentional blink task, moral and emotional words were recognized more successfully than neutral words under limited attention. Real-world tweets containing moral and emotional content that showed higher attention-capture ability in the lab were associated with more retweets (Brady, Gantman, et al., 2020).

While information that is highly moralized can draw more attention, information that is both highly moralized and incongruent with one's beliefs can potentially elicit unique attentional patterns and psychological processes. Research has shown that information that does not align with a person's political beliefs can sometimes make them feel threatened and anxious, and attract attention (Marcus & MacKuen, 1993). People are also slower in responding to information that opposes their beliefs (Lodge & Taber, 2005; Redlawsk, 2002). These findings are consistent with the broader literature on negativity bias (Baumeister et al., 2001; Rozin & Royzman, 2001), which holds that organisms tend to give greater priority to negative events, objects, and personal traits. For instance, one study by Fiske (1980) found that information about a person received greater attention (looking time) when it was negative and extreme. Moreover, a model that incorporates both information negativity and extremity accounted for likability ratings better than models that included only one of these factors. Zooming in to the belief and attitude domain, research has shown a "disconfirmation bias" (Edwards & Smith, 1996; Lord et al., 1979), that information that "disconfirms" an individual's prior opinion receives more critical evaluation and scrutiny, and this bias is more pronounced when people report stronger feelings toward the issue. In a more recent study, researchers found that thinking processes regarding information incongruent with prior political convictions are more complex and voluminous, and include more rationales (Erisen et al., 2018). Given that moralized attitudes are also experienced with heightened emotions, great personal significance, and perceived objectivity, it is reasonable

to suspect that when an opposing attitude is strongly moralized, it may draw more attention and caution during decision-making. Therefore, this chapter also examined moral misalignment, defined as the extent to which a belief-incongruent position is moralized. Specifically, its impact on response time and dwell count was tested, and the drift diffusion model (DDM) was used to characterize its more fine-grained impact on decision dynamics, particularly decision caution.

Drift diffusion models (DDM) assume that during a binary decision, stochastic evidence to choose one response (compared to the other) is accumulated over time and the decision is made when the integrated evidence reaches a decision boundary (Krajovich et al., 2010, 2012; Ratcliff, 1978; Ratcliff & McKoon, 2008). There are a few key parameters in a DDM. First, the drift rate v refers to the quality or strength of evidence accumulation. A higher drift rate toward one option usually means easier discrimination between the two options, better evidence quality of information extracted from the stimuli, and is often associated with faster responses. Second, boundary separation (a) is the amount of evidence a decision-maker needs to reach a decision. Greater boundary separation usually entails more caution and slower response time. If drift rate specifies how quickly and in which direction evidence accumulates, and boundary separation specifies how far the accumulation must travel before a decision is reached, then a complete description of the accumulation process also requires specifying where that process begins. This is captured by the starting point parameter (z), which indicates the initial position of evidence accumulation between the two decision boundaries. When the starting point is centered, the decision-maker starts unbiased between the two options, whereas when it is closer to one boundary, less evidence is needed to choose that response. In addition to the parameters that describe the evidence accumulation process, the non-decision time (T_{er}) denotes the time during the decision not spent accumulating evidence. This process can include visual encoding, stimulus

preprocessing, and motor latency. Non-decision time is important to model, as some prompts can take longer to understand and encode than others, and should not be modeled as slower evidence accumulation or greater boundary separation.

While more often applied in consumers' choices, DDM has been applied to investigate moral decisions. For example, Andrejević and colleagues (2022) found that when making moral judgments of others' behaviors, negatively valenced actions were associated with slower judgments, smaller absolute drift rates, and a biased starting point away from them. Relatedly, researchers have extended the DDM to allow momentary visual attention to modulate drift rate (Krajbich et al., 2010, 2012) and formalize it as the attentional drift diffusion model (aDDM). Although somewhat scarce, some research suggests that incorporating gaze direction into the standard DDM can improve predictions of moral decisions (Pärnamets et al., 2014). In the current study, rather than modeling the modulating effect of moment-to-moment visual attention on evidence accumulation, trial-level gaze bias was used as a predictor of drift rate. This approach was chosen because the primary goal was to test whether visual attention is associated with the direction and strength of evidence accumulation, rather than to estimate the precise temporal dynamics of fixation-contingent evidence accumulation.

Taken together, previous research suggests that moral convictions are often experienced as strongly emotional, moral and emotional content captures attention, and visual attention can both reflect and guide psychological processes during moral decisions. However, it remains less well understood how moral conviction affects moment-to-moment allocation of visual attention during socio-politically and behaviorally relevant decisions. The present study tested whether sociopolitical issues held with stronger moral conviction would also attract greater attention and more early attention. Specifically, it was hypothesized that higher moral conviction about a

sociopolitical issue predicts a higher dwell proportion to that sociopolitical issue, as well as a higher probability of the first fixation. It also examined whether moral misalignment, the extent to which a belief-incongruent position is moralized, induced more caution during decisions. It was predicted that greater moral misalignment is associated with slower decisions, greater boundary separation, and more dwell accounts.

Methods

Participants

Participants were a subset of the participants who completed the fMRI study in Chapter 2. Eye-tracking data were successfully collected for 33 of them (16 females, 16 males, 1 non-binary; age range: 18–48 years; $M_{age} = 22.70$, $SD_{age} = 5.39$).

Procedures

Procedures were described in the Chapter 2 Methods/Procedures section. Prior to the start of the decision-making task, participants completed a 9-point calibration followed by validation implemented in the EyeLink 1000 Plus Eye Tracker. Calibration and validation were repeated until acceptable accuracy was achieved. If acceptable accuracy could not be obtained within the time window that allowed the scanning session to finish on schedule, the task proceeded without eye-movement recording. All participants had normal or corrected-to-normal vision (using MRI-compatible corrective lenses when necessary).

Data Preparation

For the main analyses, AOIs were defined by splitting the screen into halves by the vertical midline. In each trial, one option (including the issue name and an image depicting protesters for or against the issue) was presented on the left half of the screen, and the other option was presented on the right half of the screen. Dwell time on either half was used as a proxy for

attention and engagement with each issue. For robustness analyses, rectangular AOIs were defined for the text and image associated with each issue using fixed screen coordinates. The left AOI was defined as $x = 70-610$, $y = 85-628$, and the right text AOI was defined as $x = 670-1210$, $y = 85-628$. These AOIs include less empty space than the left and right halves of the screen. Trials where the total dwell time or the total fixation time was shorter than 300ms were excluded from the analyses.

In the current design, participants rated their support for each sociopolitical issue presented in the decision-making task on a 7-point scale, ranging from strongly oppose to strongly support, with the mid-point being neutrality. Participants also rated their moral conviction about each issue on a 1-5 scale. These two ratings were combined to create the moral misalignment variable. For issues that participants opposed, indicated by ratings of strongly oppose, oppose, or slightly oppose, moral misalignment was computed by the moral conviction rating multiplied by 1. For issues that participants supported, indicated by ratings of strongly support, support, or slightly support, moral misalignment was calculated as moral conviction rating multiplied by -1. For neutral support issues, moral misalignment was coded as 0. For each trial, trial-level moral misalignment was defined as the maximum moral misalignment score across the two issues. In some sense, trial-level moral misalignment is analogous to a form of negativity salience, but more specifically captures the moralization of an opposed position rather than negative valence in general. In other words, high trial-level moral misalignment indicated that at least one issue in the decision was strongly moralized and opposed by the participant. Low trial-level moral misalignment suggested that neither issue was strongly morally opposed, although the trial could still include an issue that the participant strongly moralized and supported.

Data analyses

Analyses answered two series of questions. The first series concerns how moral conviction influences attention during sociopolitical decision-making. Specifically, trial-level analyses were carried out to address whether an option gets a larger share of attention or is more likely to capture early attention than its competitor when the decision-maker has a higher moral conviction in it. Complementary option-level analyses were employed to further examine whether higher moral conviction on an option attracts more absolute attention. In such option-level analyses, each trial contributed two observations, one for each issue presented. The second series focused on moral misalignment and asked whether the more an individual moralizes an opposing attitude, the more likely they are to be slower, more engaged, and cautious in making decisions.

To answer the question of how attention was divided between the two options within a decision, particularly how moral conviction affects the attention allocation while controlling for attitude strength, a mixed effects linear regression model with relative dwell time proportion as the dependent variable was fitted. For each trial, the dwell proportion for a given option was calculated by dividing the dwell time on that option by the sum of dwell times on the two options presented in that trial. The relative dwell proportion for the option on the left (dwell proportion left – dwell proportion right) was included as the dependent variable. The choice of the left option is arbitrary, and the main effect of interest is the relative characteristics between the two options. This relative dwell proportion was chosen as the primary outcome variable because the task is comparative in nature, and attention to one issue is inherently constrained by attention to the other one. Using proportion as the outcome variable also controls for total viewing time

differences across trials. Raw dwell time can partly reflect longer overall deliberation on a trial, making it a less precise index of relative attention to one issue vs. the other. Fixed effects included relative moral conviction (moral conviction left – moral conviction right), relative attitude strength (left - right), and the interaction between them. Attitude strength was operationalized by folding on the midpoint of the 7-point support rating scale, which originally ranged from strongly oppose to strongly support. This transformation led to a 0-to-3 scale, where higher values indicated stronger support or stronger opposition to the protesters' position in a given trial, and lower values indicated weaker attitudes. Since attitude strength reflects the intensity but not the direction of an attitude, it was unaffected by whether the protesters supported or opposed the issue. Support ratings were adjusted to the protesters' position. In trials with a thumbs-up icon (protesters supported the issues), support ratings were left unchanged, whereas in trials with a thumbs-down icon (protesters opposed the issues), they were multiplied by -1. This procedure ensured that the support ratings always reflected support for the *protesters'* position, although it did not affect the final attitude strength score. Main effects of whether the left option was chosen, which side of the screen that option was presented, and familiarity level with the option were also included as fixed effects. Random intercepts of participants were modeled. As expected and consistent with previous literature, there was a moderate correlation between moral conviction and attitude strength [$r(6936) = 0.49, p < .001$]. To address potential multicollinearity, both moral conviction and attitude strength were mean-centered and standardized from the raw measures. In order to maintain as much data as possible in the analyses, and none of the demographic variables were predicted to systematically affect relative dwell proportion, the main model did not include any demographic covariates. Robustness check

analyses, including participants' age, education, gender, party alignment, political engagement, and income as covariates, were conducted.

A mixed-effects logistic regression model was fitted to examine whether high moral conviction about an issue draws initial attention to it, particularly after controlling for attitude strength. The dependent variable was whether the first fixation was within the AOI of the left issue. Fixed and random effects were the same as those used in the abovementioned model, where relative dwell time proportion was the dependent variable.

For moral misalignment analyses, two mixed-effects linear regressions were conducted. Response time and total dwell count were included as the dependent variables in the two models, respectively. Support difference, moral conviction difference, trial-level moral misalignment, and the interactions between support difference and moral conviction difference, and between support difference and trial-level moral misalignment were the main independent variables of interest. Main effects of protest view and the sum of familiarity of the two issues in one trial were also included. Participant was modeled with random intercepts.

To characterize latent decision dynamics, a hierarchical attention drift diffusion model (DDM) was fit in Python using the *hssm* package. A trial-level dataset constructed for diffusion modeling was used, with responses coded as -1 and 1. Participant-level variability was accounted for using random intercepts. An outlier component was included in the model ($p_outlier = 0.05$) to account for trials that are unlikely to have been generated through the standard diffusion process. Posterior estimation was conducted using the NumPyro implementation of the No-U-Turn Sampler with four chains, 2,000 tuning iterations, and 4,000 posterior draws. The model estimated two core parameters at the participant level (indexed by j) and the trial level (indexed by i): Drift rate $v_{i,j}$ and boundary separation $a_{i,j}$. Drift rate was

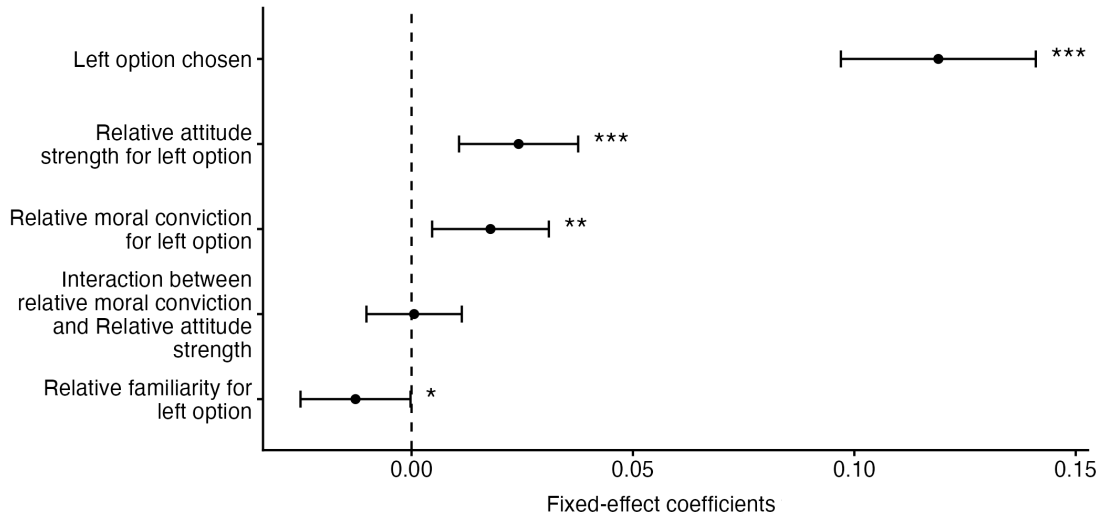
modeled as a function of support difference, moral difference, their interaction, and gaze bias proportion. This specification tested two main hypotheses: 1) whether the rate of evidence accumulation was shaped by relative support for the two options; 2) whether relative moral conviction amplifies relative support's effect on drift rate; 3) whether the degree to which visual attention favored one side over the other increases drift rate. Boundary separation was modeled as a function of moral misalignment and maximum moral salience, allowing the model to test 1) whether response caution increases as the extent to which an individual moralizes an attitude that conflicts with their beliefs increases 2) whether response caution increases as the overall moral conviction (operationalized by the maximum moral conviction of the two issues) of a decision increases. Non-decision time was estimated from the data, with participant-specific random intercepts included to account for individual differences in non-decision time.

$$\begin{aligned}
 v_{ij} &= \beta_0^{(v)} + \beta_1^{(v)} \text{SupportDiff}_{ij} + \beta_2^{(v)} \text{MoralDiff}_{ij} \\
 &\quad + \beta_3^{(v)} (\text{SupportDiff}_{ij} \times \text{MoralDiff}_{ij}) + \beta_4^{(v)} \text{GazeBiasProp}_{ij} + u_{0j}^{(v)} \\
 \log(a_{ij}) &= \beta_0^{(a)} + \beta_1^{(a)} \text{MoralMisalignment}_{ij} + \beta_2^{(a)} \text{MaxMoral}_{ij} + u_{0j}^{(a)}
 \end{aligned}$$

Results

From the model where the relative dwell proportion was the dependent variable, significant main effects of relative moral conviction [$B = 0.018$, 95% CI (0.005, 0.031), $p = .008$; Figure 3.1] and relative attitude strength [$B = 0.024$, 95% CI (0.011, 0.038), $p < .001$] were found. This suggests that both moral conviction and attitude strength increased dwell proportion, even when controlling for each other. Relative dwell proportion was significantly higher on the chosen option [$B = 0.119$, 95% CI (0.097, 0.141), $p < .001$], while it was significantly lower on the option that was rated higher in familiarity [$B = -0.013$, 95% CI (-0.025, -0.0002), $p = .046$]. The interaction between relative moral conviction and relative attitude strength was not significant,

indicating no evidence that moral conviction and attitude strength reinforce each other's effect on dwell proportion.



Results from the mixed-effects logistic regression testing the effect of moral conviction

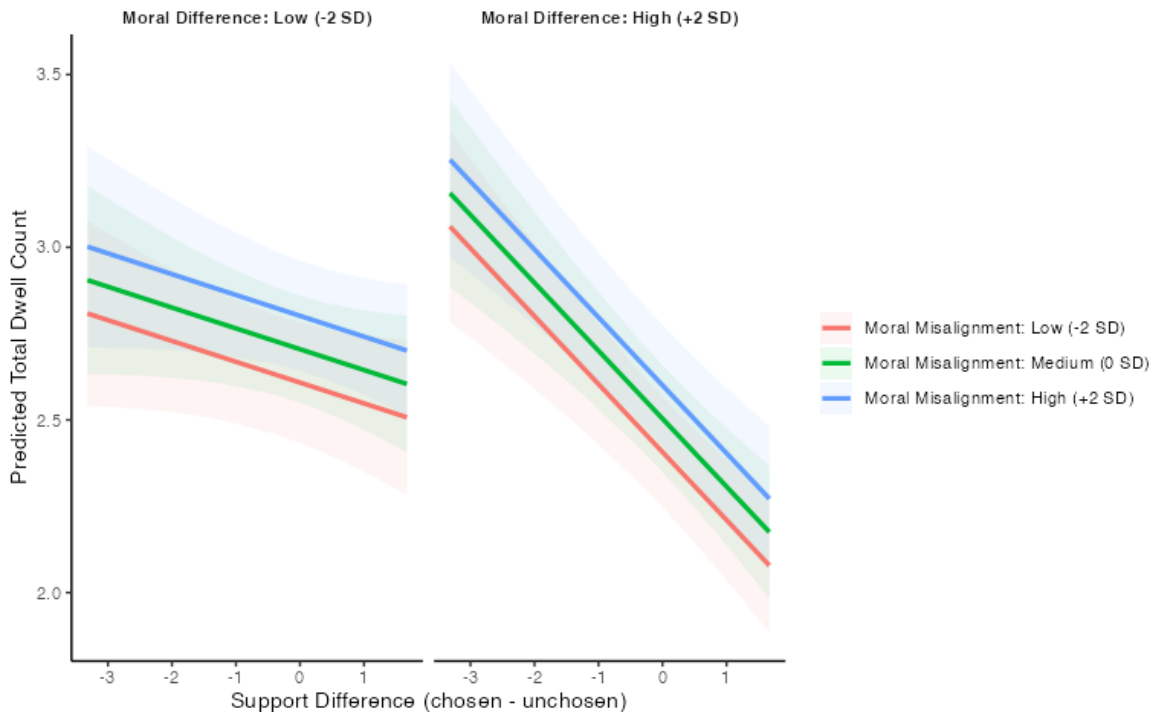
Figure 3.1: Fixed-effect coefficient plot from the model predicting the relative left dwell time proportion with 95% confidence intervals.

on first fixation showed no significant main effects of moral conviction, attitude intensity, familiarity, or whether the option was chosen. There was a significant intercept, suggesting that participants tended to first fixate on the left option when the differences in moral conviction, attitude intensity, or familiarity between the two options were at their mean levels and when the left choice was not chosen. In other words, conditional on these variables, the model predicted about an 83% probability that people would first fixate on the left choice, indicating a strong left-side bias for first fixation.

As expected, the mixed-effects linear model predicting response time revealed a significant positive main effect of trial-level moral misalignment ($\beta = 83.857$, 95% CI [62.704, 105.01], $p < .001$) that participants slowed down when they encountered an opposing attitude they moralized to a greater extent. Meanwhile, results showed significant negative main effects of familiarity ($\beta = -66.47$, 95% CI [-88.979, -43.962], $p < .001$), moral difference ($\beta = -37.329$,

95% CI [-57.266, -17.392], $p < .001$) and support difference ($\beta = -120.93$, 95% CI [-141.578, -100.282], $p < .001$), that participants responded faster on trials involving more familiar issues, when there was a larger difference in the perceived moral significance of the chosen vs. the unchosen issue, and when the chosen issue was supported to a greater extent than the unchosen issue. Whether the protesters supported or opposed the issues also had a significant impact on response time ($\beta = -300.051$, 95% CI [-338.326, -261.776], $p < .001$). The interaction between moral difference and support difference was not significant ($\beta = -11.968$, 95% CI [-32.47, 8.534], $p = 0.253$), indicating that the moral advantage of the chosen option and the support advantage of the chosen option each contributed to faster responses, but they did so additively rather than synergistically in this model.

The mixed-effects linear model testing total dwell count showed significant positive main effects of moral misalignment ($\beta = 0.048$, 95% CI [0.017, 0.08], $p = .003$; Figure 3.2) that when a decision involves at least one opposing attitude, the decision-maker moralizes to a greater extent, the more times they dwell on that decision, suggesting potential for more engagement and scrutiny. Support difference ($\beta = -0.128$, 95% CI [-0.159, -0.098], $p < .001$) and moral difference ($\beta = -0.05$, 95% CI [-0.08, -0.02], $p < 0.001$) were associated with fewer dwells. Participants made more dwells when the protesters supported the issue in question (thumbs-up) than when they opposed it (thumbs-down) ($\beta = -0.106$, 95% CI [-0.163, -0.049], $p < .001$). The interaction between support difference and moral difference was significant and negative ($\beta = -0.034$, 95% CI [-0.065, -0.003], $p = .029$), indicating that the dwell-reducing effect of support difference becomes stronger when the chosen option also holds a larger moral relevance advantage.



The drift diffusion model (DDM) parameters converged successfully, as all parameters
 Figure 3.2: Effects of support difference, moral difference and moral misalignment on total dwell count in a trial.

showed $\hat{R} = 1$ and high effective sample sizes, indicating reliable model estimation. As shown in Table 3.1, support difference (chosen - unchosen) showed a positive association with drift rate ($\beta = 0.75$, $SE = 0.02$, 95% CI [0.72, 0.79]), indicating that greater differences in support between options were associated with faster evidence accumulation toward the preferred choice. Gaze bias proportion also positively predicted drift rate ($\beta = 0.10$, $SE = 0.03$, 95% CI [0.05, 0.16]), suggesting that increased visual allocation toward an option facilitated evidence accumulation toward that decision. Although this effect was expected, it provides a useful internal consistency check that the model captures the contribution of attentional allocation to the choice process. Neither moral conviction difference (chosen - unchosen) nor its interaction with support difference was reliably associated with drift rate. Thus, the current results do not support the possibility that greater moral conviction for the chosen option increased drift rate toward that

option. Nor do they suggest that the effect of support difference on evidence accumulation reliably varied as a function of moral conviction difference. Moral misalignment showed a reliable positive association in predicting boundary separation ($\beta = 0.079$, $SE = 0.008$, 95% HDI [0.063, 0.094]), indicating that decisions were taken more cautiously when the trial included a more strongly morally opposed issue. In other words, greater moral misalignment was associated with greater caution in responses.

Table 3.1: Drift diffusion model estimates for support difference, moral misalignment, and gaze bias. The table summarizes key fixed effects from the DDM and the decision parameters associated with them.

Term	Parameter	β	95% HDI	Sig.
Moral Misalignment	Boundary Separation	0.079	[0.063, 0.094]	***
Gaze Bias Proportion	Drift Rate	0.104	[0.051, 0.159]	***
Moral Difference	Drift Rate	-0.003	[-0.035, 0.028]	n.s.
Support Difference	Drift Rate	0.754	[0.716, 0.791]	***
Support Difference $\times$ Moral Difference	Drift Rate	0.021	[-0.015, 0.056]	n.s.

To visualize these effects, simulated drift trajectories based on the model estimates are shown in Figure 3.3. The simulation illustrates how a greater support difference increases the rate of evidence accumulation, whereas a greater moral misalignment increases decision thresholds, leading to slower commitment to a choice. Together, these findings suggest that preference strength primarily influenced the speed of evidence accumulation, whereas the moralization of an opposing attitude primarily influenced decision caution.

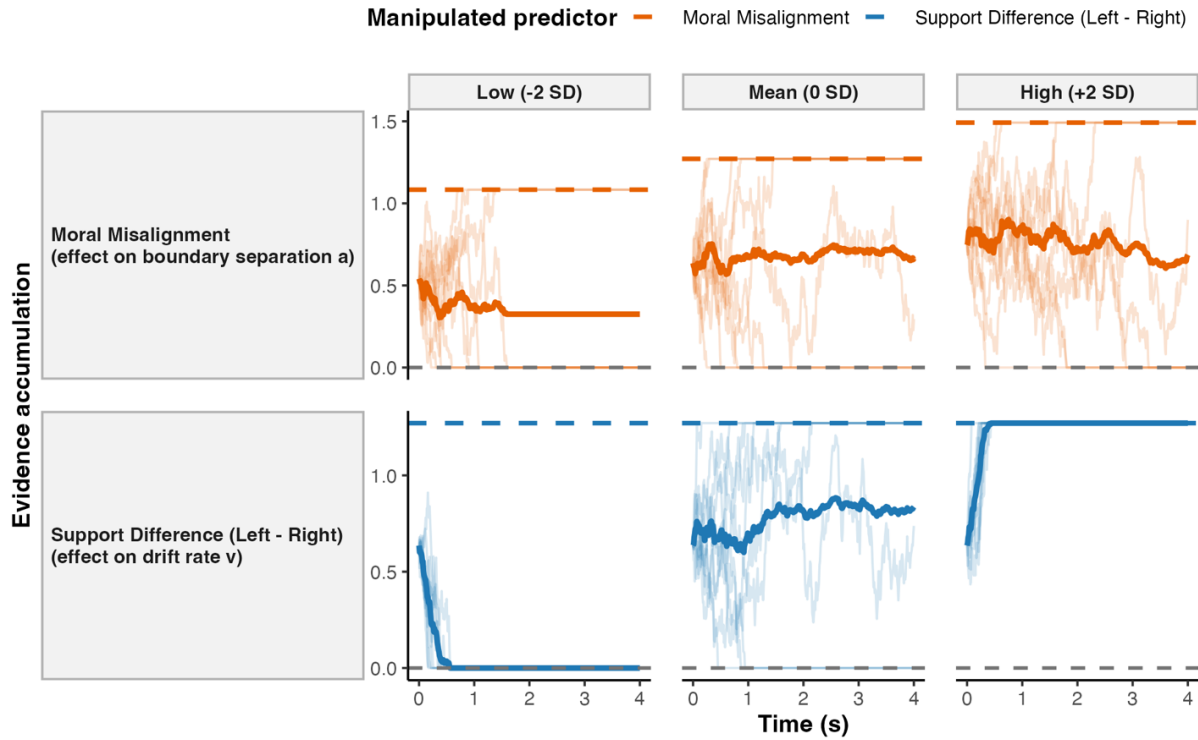


Figure 3.3: Illustrative DDM trajectories showing separate effects on drift rate (v) and boundary separation (a).

Discussion

Moral conviction is a double-edged sword. It attracts attention and supports activism but can also fuel polarization and dogmatism. This study moves beyond the behavioral impact of moral conviction and examines its role in attention allocation and decision-making mechanisms in sociopolitical contexts. Results suggest that both higher relative moral conviction and higher relative attitude intensity of an option in a decision are associated with a greater relative dwell proportion, indicating that both the congruency strength of an attitude and the extent to which the attitude is moralized attract more attention during decision-making. This chapter also examined the impact of moral misalignment on decision dynamics and found that it is associated with slower responses and greater caution during decision-making. Together, these results offer a coherent picture of the attentional dynamics associated with moral conviction. This work

complements existing work on moral conviction, which has primarily focused on downstream outcomes, by offering a more proximal cognitive understanding of how morally charged information attracts attention and affects caution during sociopolitical decisions.

While previous research and results presented in Chapter 2 support that greater overall moral conviction (operationalized by the maximum moral conviction of the two issues in a decision) is associated with faster responses, current results show that greater trial-level moral misalignment (operationalized as the maximum signed moral conviction across the two options, with moral conviction coded as positive for opposed issues and negative for supported issues) is associated with slower responses. In other words, higher moral conviction for a stance that conflicts with an individual's own position is associated with slower decisions. Moreover, moral misalignment is associated with higher total dwell counts in that trial, suggesting greater engagement and potentially more critical evaluation, consistent with previous research showing the general effect of negativity and negative valence (Fiske, 1980; Rozin & Royzman, 2001). DDM results further illustrated that greater moral misalignment may lead to larger boundary separation, demonstrating that when making decisions about a topic that goes against one's moral conviction, people may require more evidence to decide and become more cautious (Ratcliff & McKoon, 2008). Larger boundary separation often results in fewer errors (Forstmann et al., 2016). While in the current design, there are no objectively correct choices, one way to test this is to investigate whether individuals' choices are more likely to be consistent with their prior attitudes when the moral misalignment of a trial is larger. Together, to the author's knowledge, the current study is the first to systematically examine the effects of intense moralization of an incongruent attitude on sociopolitical decisions and makes a novel contribution to establishing

that the negative moralization, or moralization of a stance that one opposes, may have some unique effect on decision-making processes.

Meanwhile, little evidence supports the hypothesis that higher moral conviction attracts first fixation. This is potentially due to the complexity of the current stimuli and task structure. Despite that, there is evidence suggesting that emotion intensity may lead to early attention. First fixation is also influenced by many other factors, such as reading direction and bottom-up salience. As shown in the current results, people predominantly fixated on the left half of the screen first, which is consistent with the fact that the issue text was the major source of information and the participants were English readers who read from left to right.

Beyond situational factors, individual differences in moral attentiveness exist, suggesting that people differ in the extent to which they chronically perceive and consider the moral aspects of their experiences (Reynolds, 2008). Higher moral attentiveness has been associated with overrepresenting the moral aspects of experience, greater moral awareness, and morally good behaviors. Future research with greater statistical power should examine how this individual difference interacts with specific decision behaviors and attentional patterns in contexts involving moralized topics, thereby providing a fuller understanding of how situational and trait-level factors affect sociopolitical decisions.

CHAPTER 4

MORAL CONVICTION AFFECTS FACTUAL JUDGMENTS AND METACOGNITIVE ABILITIES

Emerging work suggests that general metacognitive abilities are associated with socio-politically consequential behaviors, while little is known about how metacognition operates in moralized vs. non-moral contexts. This chapter extends research on metacognition into consequential socio-political domains by introducing a novel semantic knowledge task that assesses metacognitive sensitivity across different levels of moral conviction. The results showed that participants' confidence ratings were higher for high-moral-conviction items, even though their accuracy in first-order factual judgments for such items was lower, indicating reduced confidence-accuracy coupling as moral conviction increased. In addition, participants were more likely to judge answers aligned with their prior views as correct, and this fact-view consistency effect was more pronounced for highly moralized content. Further analyses showed that higher moral conviction generally increased confidence but did not create a corresponding metacognitive advantage to metacognition when the correct answer aligned with participants' prior views. Instead, metacognitive sensitivity was especially reduced when correct answers conflicted with participants' highly moralized attitudes. In summary, these findings provide direct evidence that moral conviction is not only about attitudes or behavior but also shapes the second-order evaluations of one's own judgments, identifying one pathway through which moralized beliefs may distort engagement with factual information.

Introduction

People not only disagree about values, but also about empirical claims and preferences. Factual disagreement can sustain polarization because it blocks shared evidence and the common ground for compromises. Researchers on political polarization have proposed the concept of *factual belief polarization* (in addition to *ideological* and *affective polarization*) to describe partisan divides in perceptions of reality (Lee et al., 2021; Rekker, 2022). For instance, during the COVID-19 pandemic, Republicans and Democrats held different beliefs about the effectiveness of social distancing and the future severity of the pandemic (Allcott et al., 2020); they also differ in beliefs about whether climate change is happening, whether it is human-caused, and how serious it is (Dunlap et al., 2016). Outside of the political domain, disagreement about facts can emerge in many contexts, regardless of whether the underlying scientific evidence is relatively clear or genuinely mixed. From a nationally representative sample of 1,134 U.S. adults, Hamilton (2022) found that although relatively a small portion of people (ranging from 9-12%) believed in conspiracies such as NASA astronauts did not really land on the Moon, for scientific statements that exist scientific consensus, such as vaccines have mostly been a benefit to human health, the agreement ranged from 58 to 83 percent, showing non trivial disagreements or unsureness in these scientific consensus.

Motivated reasoning and confirmation bias

Why do people disagree on facts? One possibility is that they are exposed to different evidence, especially when relevant evidence is mixed, uncertain, or has changed over time. Regardless of the reasons why individuals are exposed to different evidence, such exposure may naturally lead to different conclusions. For example, people may have different opinions on whether reading or watching videos is better for learning and memory recall because conclusions are contingent on

specific contexts and conditions. While some evidence has shown that reading a transcript of a broadcast media news story leads to higher levels of memory for news content (Mesbah, 2005), other research has demonstrated that middle-school students who watched a TV science series had better word knowledge and information recall compared to those who learned the materials by reading the textbook for language acquisition purposes (Neuman & Koskinen, 1992). However, factual disagreement can persist even when people encounter similar information. One study showed that individuals with different prior beliefs can update their beliefs in opposite directions after evaluating the same evidence (Said et al., 2021). This suggests that disagreements stem not only from evidence itself, but also from internal psychological processes that integrate existing beliefs. Motivated reasoning is an obvious candidate mechanism, which states that people are motivated to believe what they want to believe (e.g., Kunda, 1990). Such motivation can affect both the selection and evaluation of information. It may influence how people recruit information, or what evidence is given consideration, as well as how they interpret and evaluate information. As a result, people with different goals or trying to preserve different prior beliefs may interpret the same information differently and come to different conclusions (Epley & Gilovich, 2016).

Various motives may drive motivated reasoning. For example, social goals such as fitting in with one's social group and maintaining social acceptance are important (Spears, 2021). To achieve these goals, people become more likely to think alike others around them, such as friends, family members, and colleagues. At a broader level, social norms and cultural expectations can fuel motivated reasoning by defining what conclusions are expected, appropriate, and desirable. Therefore, when people know that these conclusions are normative in their community, they may be motivated to interpret evidence in ways that favor or are consistent

with those conclusions (Golman, 2023; Kahan, 2013). Emotion regulation has been proposed as another reason for engaging in motivated reasoning (Thibodeau et al., 2015; Westen et al., 2006). In many cases, individuals prefer the beliefs that make them feel good and reduce negative feelings such as threat and anxiety (but see Thaler, 2026).

While motivated reasoning is a more active process during which people form or maintain beliefs based on a desired outcome rather than objective evidence, confirmation bias is a more passive process when people look for and interpret information that is consistent with their existing beliefs (Kunda, 1990; Nickerson, 1998). Confirmation bias has been found in many domains, including politics, health, and consumer behaviors (Knobloch-Westernwick et al., 2020; Meppelink et al., 2019; Yin et al., 2016). Interestingly, even in a perceptual task where participants made choices. When deciding which of the two circles contained more dots, participants spent more time reexamining the option they had initially chosen when given the opportunity to revise their decision. This biased sampling of information was associated with a higher likelihood of sticking with their original choice, even when that choice was incorrect (Kaanders et al., 2022). Taken as a whole, research on motivated reasoning and confirmation bias suggests that belief formation is not a purely evidence-driven process. Even when engaging with factual claims, people do so through the lens of their existing prior beliefs, shaped by many factors including their group memberships, cultural background, and self-identities.

Moral conviction as a force for motivated reasoning

Among the many factors that affect belief formation and drive motivated reasoning, morality plays a particularly important role. Moral rules and convictions are fundamental to the functioning of societies, as they define what is considered appropriate and what is prohibited. At the same time, moral beliefs are central to individuals' identities and social signaling (Aquino &

Reed, 2002; Westra, 2021). Because they are often experienced as objectively true and unlikely to be influenced by peers or authorities (Skitka et al., 2015), it is not surprising that they can also serve as fuel for the amplification of motivated reasoning.

For example, in a study by Cusimano and Lombrozo (2021), participants were presented with vignettes in which objective evidence supported morally unfavorable beliefs (e.g., that a friend is guilty) and asked what would be the most accurate to believe based on an impartial evaluation of the evidence, versus what the protagonist in the situation ought to believe. Results showed that people's *ought* estimates favored the morally acceptable conclusion (e.g., that the friend is not guilty) more strongly than did their judgments of what is most accurate. In the same paper, follow-up studies further demonstrated that when people had a moral reason to be optimistic, they perceived optimistic (morally desirable) beliefs as more permissible and justified, and judged evidence-based (but morally undesirable) beliefs as less permissible and justified. In line with this finding, researchers have found support for the alignment between factual beliefs about the consequences of acts and evaluations of the act's inherent morality. Specifically, the more an act was judged as inherently immoral, the less likely it was seen to produce beneficial consequences and the more likely it was to induce harm. Importantly, this association between factual judgment and moral evaluation was stronger when the content had higher moral conviction (Liu & Ditto, 2013), suggesting that moral conviction may intensify motivated thinking (see Ditto et al., 2009).

While such an uncompromising and blinding nature of moral conviction can be a source of motivation for achieving greater good in some contexts, such impact can be problematic as it can also lead to rejection of science and opposition to technology that can improve health and well-being (for review see (Kovacheff et al., 2018).

Metacognition and Moral Conviction

If moral conviction amplifies motivated reasoning, then it is also possible that it not only affects first-order beliefs and factual judgments, but also second-order evaluation of those judgments, that is, metacognition. Researchers have defined metacognition as the ability to accurately evaluate one's own knowledge and judgments, which allows individuals to represent, monitor, and adjust their own cognitive processes. This ability is crucial in education, mental well-being, and social life (Stanton et al., 2021; Carcione et al., 2019; Pescetelli et al., 2016). Metacognition has been examined using many different methods; one common approach is to ask people to answer 2-alternative forced-choice questions with a factually correct answer and to rate their confidence in each judgment. Depending on the nature of these questions, metacognition can be assessed across different cognitive domains, including perception, episodic memory, semantic knowledge, and social beliefs (Mazancieux et al., 2023; Rouault et al., 2018; Dobbs et al., 2023). In these tasks, the feature of interest is not the performance itself or how accurately people answer the questions. Instead, the primary measure is metacognitive sensitivity (meta-d') and metacognitive efficiency (meta-e), which reflect how closely confidence judgments correspond to actual accuracy. Meta-d' is influenced by task performance, making comparisons most appropriate when performance levels are similar. By contrast, meta-e is designed to adjust for task performance and is less dependent on first-level performance (Fleming & Lau, 2014; Guggenmos, 2021). Higher values on these indices mean that when individuals feel confident that they are correct, they are indeed more likely to be correct, and when they believe they responded incorrectly, they are more likely to have answered incorrectly.

A body of research has viewed metacognitive abilities as an individual trait and linked them to behaviors related to moral convictions. For example, people with lower metacognitive

ability are more susceptible to polarization, the tendency for individuals to update their prior beliefs in opposing directions after considering the same evidence (Said et al., 2021). At the same time, poor metacognitive sensitivity has been linked to radical beliefs, dogmatic opinions, and consistently adhering to only one moral principle (either utilitarianism or deontology) (Osorio T. & Reyes M., 2025; Rollwage et al., 2018; Yoder & Decety, 2022; Zmigrod et al., 2020). In the case of the EU refugee crisis, political extremists show greater certainty in their judgment about their knowledge of the crisis, independent of their actual knowledge, compared to politically moderate individuals (Van Prooijen et al., 2018). Taken together, these studies provide substantial evidence connecting poor metacognition to various sociopolitical tendencies.

Notably, these studies rely on an unstated assumption that metacognitive sensitivity is a relatively stable individual characteristic that can affect a range of other cognitive and affective processes. This assumption is important because the conclusions of these studies would be difficult to sustain if metacognitive ability varied drastically and showed little correlation across contexts. Previous work on metacognition has found some evidence in support of this domain-generalty of metacognition. To be specific, perceptual metacognitive abilities are correlated across sensory modules (e.g. Samaha & Postle, 2017; for review see Rouault et al., 2018). Outside of the sensory modules, mixed evidence has been reported regarding correlations between pairs of metacognitive efficiency in the episodic memory, visual perception, and executive function domains (Mazancieux et al., 2020). These results suggest that the extent to which confidence tracks performance across various types of tasks does, at least to some extent, rely on domain-general processes. While metacognition exhibits domain-general properties, this does not preclude variation in individuals' metacognitive performance in different contexts.

Neuroscience research has identified domain-specific impairments in metacognitive efficiency following lesions or transcranial magnetic stimulation (TMS). For instance, double dissociation has been observed for perceptual and memory metacognition following lesions to or suppression of, the frontopolar cortex and precuneus (Fleming et al., 2014; Ye et al., 2018). In the education literature, it has also been found that many stimulus-level factors can affect how well students monitor and control their own thinking. Specifically, both high familiarity and high levels of interest can lead to overconfidence and inflated feelings of knowing (Metcalf et al., 1993; Senko et al., 2022). Processing fluency, or the ease of mental processing, can also influence students' understanding of how much they have learned (Finn & Tauber, 2015). These results all point to the implication that characteristics of the content individuals engage with can affect metacognition. Therefore, although it remains underexplored whether metacognitive sensitivity is affected by the extent to which an individual moralizes particular content, existing research suggests that such an effect is theoretically plausible.

A further question is whether such an effect is not only plausible but also theoretically likely. Some insight can be drawn from research linking metacognition to a variety of sociopolitical behaviors. For instance, meta-knowledge about climate change, or how much people think they know about the subject, can explain their beliefs more than their actual knowledge about climate change (Fischer & Said, 2021). Some research has also examined partisan differences in first-order false news discernment and metacognitive awareness (Dobbs et al., 2023). A more direct clue from prior research is that engaging in moral discourse about a controversial issue seems to boost people's confidence in both moral beliefs and factual judgments, even when those judgments are false (Heinzelmann et al., 2021). Together, as discussed by Fischer & Fleming, (2024), metacognition is very important in politically contested

domains, where many topics are morally charged, and yet there is limited direct evidence on whether metacognition is poorer when individuals engage with high moral conviction content. The current study focuses on how moral conviction affects people's factual judgments, their confidence in those judgments, and their metacognitive ability, indexed by confidence-accuracy coupling. To do so, content is selected to vary in moral conviction. Individuals' accuracy in factual judgment and their metacognitive evaluation of those judgments are measured and compared across levels of moral conviction. Prior work has often emphasized either individual differences in metacognitive ability or variation across cognitive domains (e.g., perceptual and memory domains). The present study not only addresses the research gap by testing whether a content attribute within a single domain – moral conviction in factual (semantic knowledge) judgments – systematically affects metacognition but also extends the study of metacognition into the moral domain. More broadly, metacognition in the perceptual and memory domain was also measured, and exploratory analyses testing the correlation of metacognitive efficiency across domains (perceptual, memory, high moral conviction semantic knowledge, and low moral conviction semantic knowledge) were conducted. Results speak to ongoing debates about domain-general vs. domain-specific properties of metacognition.

Current Study

The current study tests two sets of hypotheses. The first set concerns the first-level performance, examining the effects of fact-view consistency and moral conviction on the accuracy of factual judgment. It is predicted that fact-view consistency leads to higher accuracy, reflecting motivated reasoning. Specifically, regardless of whether the correct answer is consistent with a participant's prior stance, they will be more likely to choose the consistent option as the correct answer. Therefore, when the correct answer favors the participant's prior stance, they will have a higher

accuracy in factual judgments, whereas when the correct answer favors the opposite side of the participant's stance, they will show a lower accuracy. The second set focuses on metacognitive abilities. First, fact-view consistency is hypothesized to inflate metacognition. Under motivated reasoning, participants are expected to report higher confidence when they choose a belief-consistent option (the option aligned with their prior stance) than when selecting a belief-inconsistent option. When the correct answer is belief-consistent, this confidence advantage would improve confidence-accuracy coupling (i.e., confidence more strongly differentiates correct from incorrect responses). In contrast, when the correct answer is inconsistent with the participant's prior stance, the same confidence advantage for belief-consistent responding would impair confidence-accuracy coupling, because high confidence would more often be associated with incorrect, belief-consistent responses. Second, two distinctive (but not mutually exclusive) hypotheses about how moral conviction affects metacognition are proposed. Hypothesis 1 predicts that moral conviction broadly disrupts metacognitive abilities and that the confidence-accuracy coupling is reduced for high-moral-conviction topics. Hypothesis 2 predicts a moderating effect: moral conviction amplifies the effect of fact-view consistency on metacognition (i.e., fact-view consistency would benefit confidence-accuracy coupling more for higher-moral-conviction topics). Importantly, these hypotheses are not mutually exclusive, that moral conviction could reduce metacognitive ability overall while also amplifying the effect of fact-view consistency.

Methods

Participants

Participants (N = 140; 69 females, 71 males; Age range: 21 –80 years; $M_{age} = 42.7$, $SD_{age} = 13.1$) were recruited through Prolific and completed the task. All participants' current country of

residence was the US, and their party and sex were balanced. All participants were compensated for completing Part 1 of the study, which collected their demographic information and their views on various sociopolitical and preferential topics. Participants who rated “not at all familiar” with topics that correspond to more than half of the questions (> 100 of 200 questions) in the knowledge task were not invited to Part 2-3, where the knowledge task was administered. All other participants were invited to Part 2-3 and were compensated an additional \$20 for completing them. Participant completed Part 1 in about 50-60 min. Parts 2 and 3 took about 35-40 min each. No part of the study procedures or analyses was pre-registered prior to the research being conducted.

Materials

Six sets of questions were prepared for estimates of participants’ metacognitive efficiency in six conditions, including a perceptual metacognition task, a memory metacognition task, and a novel 2x2 design knowledge metacognition task: topic (high vs. low moral conviction) x consistency (correct answer consistent vs. inconsistent with one’s viewpoints).

The Perceptual Metacognition Task was an adapted version of the task developed by Fleming and Lau (2014). The task was programmed using JSpsych and was run online. In each trial, participants saw two squares that contained different numbers of dots and determined which square had more dots. After each decision, participants rated their confidence in their choice on a 6-point scale. The task was self-paced. The difficulty of the task was customized for each participant during an initial practice phase (trial number = 25) in which, instead of indicating confidence, each participant received feedback (correct or incorrect) on their judgments. After this practice phase, participants completed two blocks of 25 trials for the main task.

The Memory Metacognition Task was an adapted version of the task in Mazancieux and colleagues (2020), which consisted of 2 blocks of 25 trials. Each block had an encoding phase followed by a test phase. During encoding, participants viewed 25 unrelated word pairs for 2,500 ms each in random order. Words were sampled from the SUBTLEX-US database according to the following criteria: nouns with 4-8 letters, 1-2 syllables, frequency between 20 and 100 per million. In the test phase, one word from the studied pair served as the cue, and the participant decided which one of the two options had been paired with it. The distractor options were drawn from the SUBTLEX-US database using the same criteria. After each decision, participants rated their confidence on a 6-point scale.

The Knowledge Metacognition Task was a novel metacognition task in which participants made binary choices about factual questions. In the current study, the knowledge task had 4 conditions, resulting from a 2-level moral conviction (high vs. low) x 2-level fact-view consistency (consistent vs. inconsistent) design. Each condition had 50 questions, yielding a total of 200 questions in the knowledge task. Specifically, 100 factual binary-choice questions were created for both high- and low-moral-conviction topics. Each topic had 2, 4, or 6 corresponding questions. Half of the corresponding questions had the correct answer favoring one side of the question, and the other half had the correct answer more in line with the other side. For example, in Panel A in Figure 4.1, the topic is capital punishment, and the correct answer is *a*, which is not in favor of capital punishment. The correct answer for the question in Panel B is *a*, which is more in favor of capital punishment than *b*. Panels C and D show examples of low-moral-conviction questions regarding binge-watching vs. spaced viewing of TV shows. Because the correct answers are always balanced for both sides of one topic, this design allows all participants to see 50 questions for each of the 4 conditions. All facts in the questions are supported by peer-

reviewed journals, government reports, or surveys from research centers (e.g., Pew Research). A full list of high- and low-moral-conviction questions can be found in Appendix C Appendix D. As in previous tasks, participants indicated their confidence in the correctness of their answers after each question using a 6-point scale.

A

How do the costs of life in prison compare to the costs of capital punishment, considering factors such as legal costs and time in prison?

a. Capital punishment is up to 10x more expensive than a lifetime sentence.
b. A lifetime sentence is up to 10x more expensive than capital punishment.

1 2 3 4 5

How confident are you with your judgment?

B

What is the failure rate for death by lethal injection from 1890 to 2010?

a. 7.12%
b. 15.6%

1 2 3 4 5

How confident are you with your judgment?

C

Is binge watching better or worse than spaced viewing for recall 4 weeks later of episodes from a TV series?

a. Better.
b. Worse.

1 2 3 4 5

How confident are you with your judgment?

D

Which is better for learning the vocabulary of a second-language from TV?

a. Binge watching of TV show episodes
b. Spaced viewing of TV show episodes

1 2 3 4 5

How confident are you with your judgment?

Figure 4.1: Summary of the 2x2 (high/low moral conviction x consistent/inconsistent) conditions of the novel semantic knowledge metacognition task. A, an example of the semantic knowledge task (*high moral conviction, (in)consistent* if participant opposes(supports) capital punishment), correct answer is option *a*, not in favor of capital punishment. B, an example of the semantic knowledge task, (*high moral conviction, (in)consistent* if participant supports(opposes) capital punishment), correct answer is option *a*, more in favor of capital punishment compared to option *b*. C, an example of the semantic knowledge task (*low moral conviction, (in)consistent* if participant prefers binge watching (spaced viewing), correct answer is option *a*, in favor of binge watching. D, an example of the semantic knowledge task, (*low moral conviction, (in)consistent* if participant prefers spaced viewing (binge watching), the correct answer is option *b*, not in favor of binge watching.

Procedures

The study had three parts. In Part 1, participants completed a series of ratings for all high- and low-moral-conviction topics that appeared in the knowledge task. For each topic, they rated their level of support on a 6-point scale (from strongly oppose to strongly support). Responses to this 6-point scale were then categorized as support vs. opposition, and the fact-view consistency of each question in the knowledge task was calculated. Participants also rated their level of moral

conviction and familiarity with each topic. Moral conviction for each topic was measured by the average score of two questions using a 5-point scale: “To what extent is your position on ____ a reflection of your core moral beliefs and convictions?” and “To what extent is your position on ____ connected to your beliefs about fundamental right and wrong?” (Skitka & Morgan, 2014). Familiarity with each topic was measured on a 5-point scale (from *not at all familiar* to *very familiar*). After rating all the topics, participants completed the perceptual and memory tasks. In Part 2, participants completed a random selection of 100 out of 200 questions in the knowledge task and finished the remaining 100 questions in Part 3.

Data preparation

In the knowledge task, fact-view consistency refers to the alignment between a participant’s prior stance on a topic and the stance endorsed by the correct answer to the factual question. It was coded by comparing each participant’s prior support rating on a topic with the direction implied by each question’s correct answer. The 6-point support scale for each topic was collapsed into a binary stance: the three “support” options (somewhat/support/strongly support) were labeled *Support*, and the three “oppose” options (somewhat/support/strongly oppose) were labeled *Oppose*. Each question in the knowledge task was designed to have a correct answer that either supports the topic or opposes it (e.g., a death penalty question whose correct answer indicates effectiveness was tagged *supports*, whereas one indicating ineffectiveness was tagged *opposes*). A trial was coded as consistent when a participant’s stance matched the answer direction (e.g., when the participant supports the death penalty and a question has a correct answer indicating effectiveness of the death penalty, *Support* with *supports*, or *Oppose* with *opposes*) and inconsistent when they mismatched (*Oppose* with *supports*, or *Support* with *opposes*). The same rule applied to all topics. Since, for each topic, half the questions supported the topic and the

other half opposed it, irrespective of a participant's prior stance on the topic, exactly half the questions would have belief-consistent correct answers and half would have belief-inconsistent answers.

To confirm that the predesignated high moral topics are indeed regarded as high in moral conviction by the participants, and the low moral conviction topics are regarded as low in moral conviction, participants' moral conviction ratings for the high and low moral topics were compared using a paired t-test. Results showed a significant difference in mean moral conviction ratings between the high- and low-moral topics ($t(53.75) = 19.10, p < .001$). Figure 4.2 shows the abbreviated version of the topic content and the mean and standard deviation of moral conviction on each of the topics. One item in the knowledge task was identified as invalid after data collection was completed. Both the designated correct and incorrect answers were erroneous, and this item was excluded from all analyses.



Figure 4.2: Moral conviction by topics. Ratings are shown for a subset of 140 participants (including 126 participants from the second batch), because minor changes were made to some topics and questions across batches.

Two approaches were implemented to quantify metacognition. First, in trial-level analyses, confidence-accuracy coupling indexed metacognition: larger confidence differences between correct and incorrect trials indicate better metacognition. It was tested whether moral conviction and fact-view consistency moderated this difference in confidence-accuracy coupling. Second, metacognition was quantified by metacognitive efficiency (Fleming & Lau, 2014). Accuracy and confidence were aggregated within each condition, and metacognitive efficiency ($\text{meta-d}'/\text{d}'$) was estimated using a hierarchical Bayesian signal detection model (Fleming, 2017; Maniscalco & Lau, 2012). Metacognitive efficiency was estimated for each participant separately in six conditions: high moral-consistent, high moral-inconsistent, low moral-consistent, low moral-inconsistent, perceptual, and memory. The model was implemented with the HMeta-d toolbox, which uses Markov Chain Monte Carlo (MCMC) sampling to estimate the posterior distribution over model parameters. Using JAGS (version 3.4.0) within R. Both group-level and participant-level estimates of $\text{meta-d}'/\text{d}'$ were obtained for each condition within this hierarchical framework.

Across all analyses that included demographic covariates, age was standardized. Education was converted to years of education: 12 for high school diploma; 14 for associate degree or some college, but no degree; 16 for bachelor's degree; and 18 for master's degree, professional degree, or doctoral degree. Income was coded as a continuous variable: 1 for less than \$25,000; 2 for \$25,000–\$49,000; 3 for \$50,000–\$74,999; 4 for \$75,000–\$99,999; 5 for \$100,000–\$149,999; 6 for \$150,000 or more, and participants who responded prefer not to say were treated as NA. Party alignment was treated as continuous from 0 (Strong Democrat) to 6 (Strong Republican). Political engagement ranged from 0 (never) to 4 (always), representing the extent to which participants followed politics and public affairs. In all trial-level analyses, moral

conviction was the moral conviction rating for the topic corresponding to each question standardized within participants. Confidence was also standardized within participants. Linear mixed-effects models and generalized linear mixed-effects models were fitted using the *lmer* and *glmer* function from the *lme4* package (version 1.1.37) in R (version 4.5.1 2025-06-13). Models that did not converge under the default optimizer were refitted with the *bobyqa* optimizer. Fixed-effects estimates were stable across optimizers.

Data Analyses

Task performance: How do moral conviction and prior belief affect accuracy in factual judgment?

A mixed-effects logistic regression model was fitted to trial-level data to investigate how moral conviction and fact-view consistency affect accuracy. The dependent variable was whether the participant made the correct choice (1 for correct, 0 for incorrect). Moral conviction, fact-view consistency, and their interaction were included in the model as independent variables. Random intercepts for participants and questions were included to account for individual and question-level differences. Demographic variables, including age, education, income, party alignment, and political engagement, were included as covariates.

Metacognition: How does moral conviction affect confidence-accuracy coupling?

Confidence-accuracy coupling was used as an index for metacognitive sensitivity. In the trial-level mixed-effects models in this section, confidence (centered within participant) was the dependent variable, and trial-level accuracy (correct vs. incorrect) was included as a fixed effect. The main effect of accuracy captured the extent to which confidence differed between correct and incorrect trials and served as the measure of metacognitive sensitivity.

To examine whether moral conviction regarding the topic of each factual judgment reduces confidence-accuracy coupling, a mixed-effects linear regression model was fitted to the trial-level data. Main effects of moral conviction, accuracy, and the interaction between them were included as independent variables. The dependent variable is the confidence rating in each trial. Moral conviction and confidence rating were mean-centered within each participant, and this is true for all trial-level analyses that included moral conviction and confidence. Random intercepts were modeled for participants and questions. Demographic variables, including age, education, income, party alignment, and political engagement, were included as covariates. The interaction term tests whether the confidence difference between correct and incorrect trials varies with moral conviction, corresponding to the hypothesis that moral conviction broadly disrupts metacognitive abilities, or, more precisely, that confidence-accuracy coupling is reduced for high-moral-conviction topics.

Metacognition: How do moral conviction and fact-view consistency affect confidence-accuracy coupling?

To test whether fact-view consistency increases confidence-accuracy coupling and whether moral conviction strengthens the benefit of fact-view consistency for confidence-accuracy coupling, a mixed-effects linear regression model was fitted to trial-level data. In this model, the main effects of moral conviction, accuracy, and whether the correct answer (fact) was consistent with the participant's prior view influence confidence were included as fixed effects. All possible two-way and three-way interactions were included as fixed effects as well. The dependent variable is confidence. Random intercepts were modeled for participants and questions. Age, education, income, party alignment, and political engagement were included as covariates. The two-way interaction between fact-view consistency and accuracy examines whether fact-view

consistency benefits the confidence-accuracy coupling. Since moral conviction was centered, this two-way interaction reflects the effect of fact-view consistency on confidence-accuracy coupling at average moral conviction; specifically, whether the confidence-accuracy coupling is stronger when the correct answer is consistent vs. inconsistent with the participant's prior stance. The three-way interaction among fact-view consistency, accuracy, and moral conviction tests whether moral conviction moderates this "consistency advantage" in confidence-accuracy coupling, particularly whether the difference in metacognition (indexed by confidence-accuracy coupling) between belief-consistent and belief-inconsistent trials is larger at higher moral conviction.

Metacognitive efficiency: How do moral conviction and fact-view consistency shape metacognitive efficiency?

To examine how moral conviction (high vs. low, a categorical variable) and fact-view consistency affect metacognitive efficiency, a mixed-effect linear regression model was fitted. Metacognitive efficiency for each participant in each of the four conditions in the knowledge task was included as the dependent variable. Moral conviction, fact-view consistency, and their interaction were included as the fixed effects, and a random intercept was included for each participant. Age, education, income, party alignment, and political engagement were included as covariates.

Metacognitive efficiency: Are metacognitive efficiencies correlated across domains?

Each participant's metacognitive efficiency was estimated using a hierarchical Bayesian signal detection model (Fleming, 2017; Maniscalco & Lau, 2012) across the perceptual, memory, low-moral-conviction, and high-moral-conviction domains. Pair-wise Pearson correlations were conducted to examine whether metacognitive efficiency was correlated across domains.

Results

Task performance: How do moral conviction and prior belief affect accuracy in factual judgment?

Results showed a positive main effect of fact-view consistency on judgment accuracy, suggesting that at a participant's average moral conviction, they were more likely to choose the correct answer when the correct answer is consistent with their prior belief ($p < .001$, $OR = 1.62$, $95\% CI = [1.53, 1.72]$). In other words, participants were more likely to choose the answer that is consistent with their prior belief. There was also a significant main effect of moral conviction on judgment accuracy ($p < .001$, $OR = 0.919$, $95\% CI = [0.890, 0.948]$), suggesting that at the reference level, i.e., in trials where the correct answer is inconsistent with a participant's prior view, higher moral conviction led to a lower chance of being correct. Importantly, the interaction between moral conviction and fact-view consistency was significant ($p < .001$, $OR = 1.12$, $95\% CI = [1.08, 1.17]$, see Figure 4.3), meaning that on fact-view inconsistent trials, higher moral conviction was associated with lower odds of being correct ($OR = 0.919$) and on fact-view consistent trials, higher moral conviction was associated with a slightly higher odds ($OR = 1.03$) of being correct.

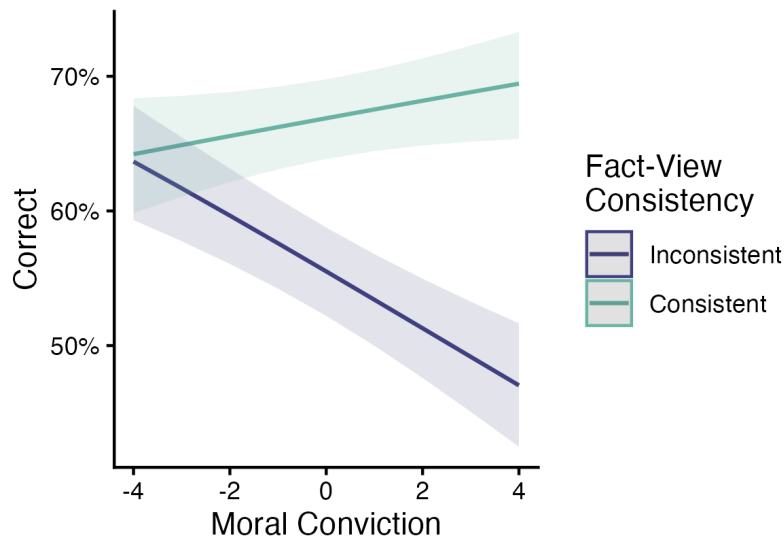


Figure 4.3: Effects of moral conviction and fact-view consistency on accuracy. The shades illustrate 95% confidence intervals.

Metacognition: How do moral conviction and accuracy affect confidence?

There was a significant positive association between moral conviction and confidence [$B = 0.054$, 95% CI (0.034, 0.075), $p < .001$; Figure 4.4], indicating that participants were more confident about their judgment when the question was about topics with higher moral conviction at the reference level (i.e., when participants chose the wrong answer). Being correct was associated with higher confidence [$B = 0.259$, 95% CI (0.225, 0.292), $p < .001$] at a participant's average moral conviction. Importantly, there was a significant interaction between moral conviction and trial-level accuracy (correct vs. incorrect) [$B = -0.026$, 95% CI (-0.050, -0.003), $p = .030$]. This suggests that the positive association between moral conviction and confidence was stronger on incorrect trials than on correct trials. In other words, moral conviction increased confidence more when participants were wrong, suggesting worse metacognitive sensitivity when engaging with higher-moral-conviction topics.

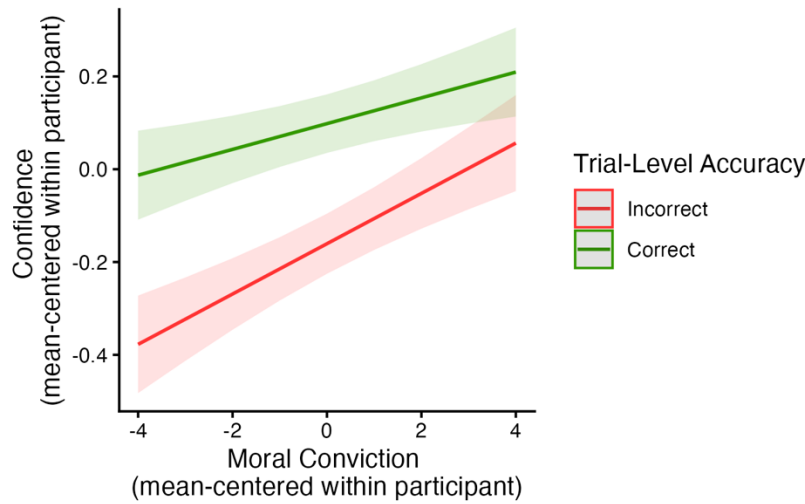


Figure 4.4: Effects of moral conviction and accuracy on confidence. The shades illustrate 95% confidence intervals.

Metacognition: How do moral conviction and fact-view consistency affect confidence-accuracy coupling?

As in the previous model without fact-view consistency, the model where fact-view consistency was added showed a significant main effect of correctness [$B = 0.088$, 95% CI (0.042, 0.134), $p < .001$; Figure 4.5] that in the reference condition (fact-view inconsistent trials), confidence is higher on correct than incorrect trials, suggesting positive confidence-accuracy coupling even when the correct answer conflicts a participant's prior belief. There was also a significant two-way interaction between fact-view consistency and correctness [$B = 0.349$, 95% CI (0.284, 0.414), $p < .001$], indicating that the confidence difference in correct and incorrect responses was larger in fact-view consistent trials. In other words, participants had higher confidence in their correctness when they chose answers that aligned with their prior beliefs. The three-way interaction among moral conviction, fact-view consistency, and accuracy on confidence tested whether this “prior-belief advantage” in confidence-accuracy coupling was stronger for more moralized topics and was significant [$B = 0.074$, 95% CI (0.027, 0.122), $p = .002$]. This three-

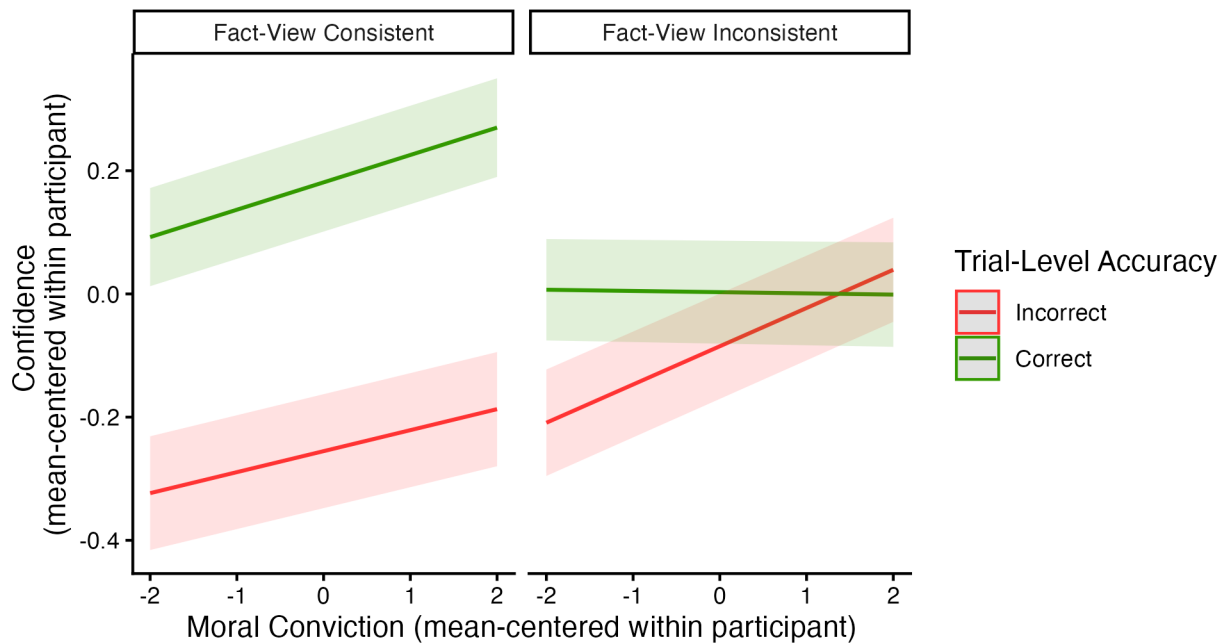


Figure 4.5: Effects of fact-view consistency, accuracy, and moral conviction on confidence. The shades illustrate 95% confidence intervals.

way interaction suggests that as moral conviction increased, the confidence gap between correct and incorrect responses shrank more on fact-view-inconsistent than on fact-view-consistent trials, and selectively undermined confidence-accuracy coupling for fact-view-inconsistent trials. More explicitly, when the correct answer conflicts with a participant's prior belief, higher moral conviction was associated with elevated confidence on incorrect (but belief-consistent) responses, making confidence less diagnostic of correctness. When the correct answer aligned with a participant's prior belief, higher moral conviction had a much weaker effect on the confidence difference between correct and incorrect responses. Together, these results provide partial support for Hypothesis 2, which posits that moral conviction moderates the effect of fact-view consistency on metacognition. However, rather than strengthening confidence-accuracy coupling when fact and view are consistent *and* weakening it when they are inconsistent, moral conviction seems to primarily amplify the latter effect. Potential explanations for this asymmetry

are discussed in the Discussion. Significant main effects of moral conviction [$B = 0.062$, 95% CI (0.036, 0.088), $p < .001$] and fact-view consistency [$B = -0.170$, 95% CI (-0.223, -0.118), $p < .001$] were found, but these effects are conditional on the reference levels of the interacting variables and should be interpreted with caution.

Metacognitive efficiency: Are metacognitive efficiencies correlated across domains?

A significant weak positive correlation between meta-e in the memory domain and meta-e in the low moral conviction domain was found $r(140) = .18$, $p = .03$. There was also a significant weak positive correlation between meta-e in the low and high moral conviction tasks $r(140) = .18$, $p = .04$. No significant correlation was found for any of the other possible pairs across the perceptual, memory, low moral conviction and high moral conviction domains.

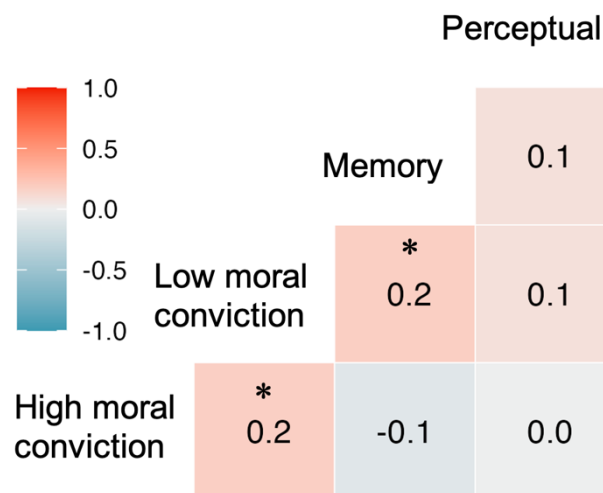


Figure 4.6: Correlation matrix of metacognitive efficiency in the perceptual, memory, low moral conviction, and high moral conviction domains. Heatmap colors represent strength of association.

Discussion

In a novel knowledge metacognition task, participants' factual judgment and metacognition were measured to examine how moral conviction and fact-view consistency affected them. The results supported both major hypotheses about moral conviction. First, higher moral conviction of the content engaged was associated with poorer metacognition, indexed by trial-level confidence-accuracy coupling and by condition-level metacognitive efficiency estimated using a hierarchical Bayesian signal detection model (Fleming, 2017; Maniscalco & Lau, 2012). Second, the association between moral conviction and metacognition depended on fact-view consistency: higher moral conviction reduced confidence-accuracy coupling more strongly in fact-view-inconsistent trials. To break it down further, for low-moral-conviction topics, confidence remained higher for correct than for incorrect responses, even when the truth conflicted with prior beliefs. For high-moral-conviction topics, this confidence gap between correct and incorrect judgments was attenuated in fact-view-inconsistent trials because confidence remained relatively high even when participants chose the incorrect (belief-consistent) option.

The results showing that individuals were more confident in factual judgments about high moral conviction topics regardless of their actual correctness is consistent with prior research suggesting that attitudes held with strong moral conviction are experienced with high certainty and as objectively true (Skitka et al., 2005, 2021). Relatedly, Heinzelmann and Tran (2024) examined beliefs about genetic technologies have shown that attitude extremity is positively associated with metacognitive certainty about those attitudes. While prior work has shown that moral conviction has a behavioral impact beyond attitude extremity, they are often correlated and may reinforce each other. One study has shown that engaging in moral discourse appears not only to boost confidence in one's moral beliefs but also to increase the extremity of those views

(Heinzelmann et al., 2021). This disproportionately high certainty in more extreme first-order beliefs may contribute to social polarization as well as moralism, a problematic way of moralization that oversimplifies, distorts, or hinders a fuller understanding of the situation (Heinzelmann & Tran, 2024; Taylor, 2012). In a study in which attitude change was formally measured, Yoder & Decety (2022) showed that beliefs held with greater moral conviction are less susceptible to social influences, suggesting that individuals may, in general, feel more confident in their moralized beliefs. The current study further contributes to this literature by showing that the confidence-enhancing effects of moral conviction extend beyond confidence in one's moralized beliefs themselves to confidence in factual judgments related to those beliefs. This distinction is meaningful because, in principle, factual judgments should be evidence-based rather than guided by subjective opinions, while the findings suggest that moral conviction may spill over from certainty of beliefs themselves to epistemic judgments related to those beliefs, biasing how people assess and use their knowledge when it comes to moralized content. Importantly, further analyses showed that this confidence-enhancing effect of moral conviction is especially pronounced when participants made factually incorrect choices. These results further illustrate how increased confidence in judgments about high moral conviction content can be concerning for society, potentially leading people to persist in incorrect perceptions of reality and causing factual disagreements between individuals who hold different moral convictions, thereby undermining the common ground necessary for discussion and solutions.

This asymmetry in the confidence-boosting effect of moral conviction on correct vs. incorrect judgments poses a caveat to the hypothesis that assumes symmetry. The hypothesis predicted that moral conviction would both (a) increase confidence in belief-consistent responses by making them feel more correct and (b) decrease confidence in belief-inconsistent responses

by making them feel more wrong. If both happen, moral conviction should widen the correct-incorrect confidence gap when the correct answer is belief-consistent (fact-view consistent trials) and shrink it when the correct answer is belief-inconsistent (fact-view inconsistent trials), leading to a three-way interaction among moral conviction, fact-view consistency, and accuracy on confidence. Although the observed three-way interaction was significant and in the expected direction, the effect appears to be driven primarily by confidence inflation for factually incorrect, belief-consistent responses and the absence of confidence inflation for factually correct, belief-inconsistent responses on high moral conviction trials. The data did not support the hypothesized additional confidence enhancement for factually correct, belief-consistent responses beyond the general confidence-boosting effect of high moral conviction. Potential psychological mechanisms underlying this finding are discussed below.

From the results, moral conviction appeared to boost confidence across most contexts, except when participants selected the correct answer on fact-view inconsistent trials. In other words, even when participants successfully inhibited the tendency to choose the belief-consistent response and arrived at the objectively correct answer, their confidence was not enhanced. One possible explanation is that, in these cases, participants may have extracted sufficient evidence to identify the correct response. However, because that response conflicted with their strongly moralized prior beliefs, the decision felt subjectively uneasy or unsmooth. This residual sense of conflict may have counteracted the general confidence-enhancing effect of higher moral conviction. One may then ask why this dampening effect of choosing a fact-view-inconsistent option did not appear symmetrically in fact-view-consistent trials. If it had, then, together with the especially increased confidence observed when participants correctly chose the fact-view-consistent option under high moral conviction, it would have resulted in enhanced metacognitive

sensitivity in the high moral conviction, fact-view-consistent condition. So, when participants incorrectly chose the fact-view-inconsistent option in high-moral-conviction, fact-view-consistent trials, they may not experience a genuine conflict between the evidence and their prior moralized belief. Instead, they may simply have formed an inaccurate impression that the belief-inconsistent option was plausible and entered a state of committed choice without strong evidence favoring it. If so, participants would feel less evidence-based unease to make the choice and thus show less corresponding suppression of the confidence-enhancing effect of high moral conviction. Some previous research indirectly supports these speculations by showing that when individuals' intuitions conflict with evidence-based reasoning, they show a decrease in confidence (De Neys et al., 2011), yet more studies are needed to both replicate the asymmetry and examine the underlying psychological mechanisms.

Although modeling each question with a random effect offers relatively clear evidence of overconfidence and poorer metacognitive sensitivity when engaging with high moral conviction content, one limitation lies in the comparison of metacognitive abilities across domains. Metacognition literature has suggested that first-level performance in a metacognition task can influence the estimation of metacognition, even for measurements such as the metacognitive efficiency, which had been designed to remove the effect of first-level performance (Guggenmos, 2021). In the current study, first-order performance in the high moral conviction knowledge task was generally poorer than in the low moral conviction knowledge task, which can lead to biased estimates of metacognitive efficiency using the hierarchical Bayesian procedures developed by Fleming and Lau (2014). Therefore, the absence of correlation across the perceptual, memory, and knowledge domains (high moral vs. low moral by fact-view consistent vs. fact-view inconsistent) is insufficient to suggest that metacognition is domain-

specific. A future direction would be to match first-order task performance across high- and low-moral-conviction knowledge tasks, as well as across metacognitive tasks in other domains, such as perception and memory. Doing so would provide a cleaner and more comparable estimation of metacognitive sensitivity and efficiency across domains.

Nevertheless, this approach should be interpreted with an important caveat that moral conviction may itself alter first-order performance. In the current study, the lower accuracy when answering factual questions about high moral conviction topics than when answering questions about low moral conviction topics may arise from two sources. First, the high-moral-conviction questions may be intrinsically more difficult. If this were the case, matching the difficulty level across high vs. low moral conviction domains by pretesting item-level difficulty in an independent sample would help isolate and identify the complete effect of moral conviction on metacognitive abilities. Second, however, high moral conviction may disturb the first-order cognitive processes involved in judgment formation, such as evidence evaluation and selection (see Liu & Ditto, 2013). In this case, matching accuracy across tasks should be interpreted with caution, as a test of whether observed metacognitive differences are reducible to first-order performance differences, rather than as an estimate of the full effect of moral conviction on metacognition. In other words, moral conviction may affect metacognition through both a direct effect on second-order monitoring and an indirect effect mediated by disruptions to first-order judgment processes. Notably, in the primary analyses of the current study, questions were modeled as random effects, which helped account for baseline differences in item difficulty. Thus, the observed effect of moral conviction on confidence-accuracy coupling is unlikely to be explained solely by differences in question difficulty. Together with the first-order results that moral conviction was associated with disruptions in first-order factual judgment, these findings

provide evidence consistent with the dual-pathway account. Yet the present study was not designed to decompose the relative contribution of direct effects on second-order monitoring or indirect effects through first-order judgment processes. Future research should more directly test and quantify this dual-pathway account.

Appendix B

Political party alignment moderates the effects of moral conviction and fact-view consistency on accuracy in factual judgments

As an exploratory analysis, a mixed-effects linear regression model was fitted to test whether party alignment (from 0 = Strong Democrat to 6 = Strong Republican) moderates the effects of moral conviction and prior belief on accuracy in factual judgments. Party alignment, moral conviction, fact-view consistency, and all possible two-way and three-way interactions among them were modeled as independent variables. Random intercepts for participants and questions were included. Results showed a significant three-way interaction among moral conviction, fact-view consistency, and party alignment, suggesting that the amplification of motivated reasoning by moral conviction is stronger among Democrats than among Republicans.

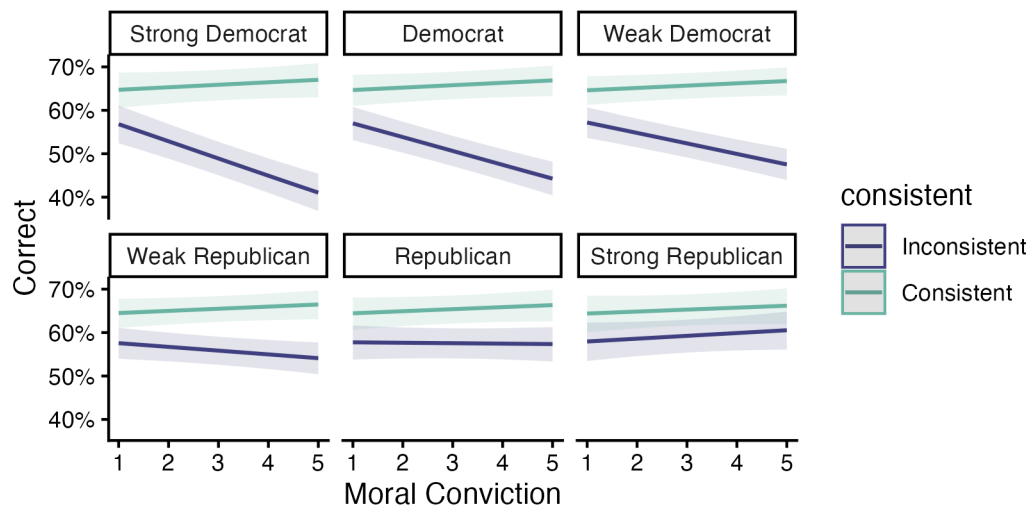


Figure 4.7: Effects of fact-view consistency, moral conviction, and party alignment on factual judgment correctness. The shades illustrate 95% confidence intervals.

CHAPTER 5

GENERAL DISCUSSION

Using behavioral experiments, eye-tracking, and fMRI, this dissertation addresses the overarching question of how moral conviction shapes the way people attend to, respond to, and evaluate sociopolitical information and decisions. It extends existing work that primarily focuses on the functions of moral conviction. Beyond the established role of moral conviction in motivating collective action, increasing political participation, and prompting avoidance of those with opposing views, evidence across studies suggests that it also influences the psychological processes by which people interpret and engage with the world. Accordingly, this dissertation examines the effects of moral conviction on both first-order decision processes and second-order metacognitive processes.

For first-order processes, during decision-making about sociopolitical issues, the extent to which individuals moralize those issues may not directly affect the decisions themselves, which are predicted primarily by supportive or opposing attitudes. Moral conviction, however, attracts greater attention and accelerates decision-making, even when controlling for attitude strength, suggesting that it affects the psychological processes through which people arrive at their conclusions. Relatedly, the fMRI results converge with theoretical accounts of both moralization and moral conviction in suggesting that these constructs involve both cognitive and affective dimensions (Feinberg et al., 2019; Wright et al., 2008). Particularly, higher moral conviction is associated with greater hemodynamic responses in the IPFC, a region that has been associated with executive control and goal-directed behaviors (see Carlson & Crockett, 2018; Dixon & Christoff, 2014), as well as in salience network, including the aINS and ACC, which have been linked to affective processing (see Sevinc et al., 2017; Uddin, 2015). Interestingly, fMRI results

also showed elevated functional connectivity between lPFC and vmPFC/mPFC during decisions involving sociopolitical issues rated as having higher moral conviction, suggesting a possible integration of moral values into sociopolitical considerations.

Chapter 4 examined a different type of first-order cognitive process and found that during information processing and factual judgments, moral conviction seems to accentuate motivated reasoning, pushing people to judge facts that are aligned with their moral convictions as true more often than they judge those that do not align with their convictions, regardless of the actual truthfulness. This has societal implications that, even for factual judgments that should be evidence-driven and have correct answers, when it comes to topics that are greatly moralized, people are affected by their opinions rather than basing their judgments on facts. As recent studies have shown, over the past few decades, people have become more likely to moralize certain concepts across diverse categories, and to use moralized language in both newspapers and social media discourse (Puryear et al., 2025; Ramezani et al., 2026). These co-occurred with the trend of US society becoming increasingly polarized (Boxell et al., 2024). Results from this dissertation provide a mechanistic account that helps link these trends, such that as moralization increases, it becomes harder for people holding different views to agree on factual common grounds, which in turn, may lead to misunderstanding and more dislike of the disagreeing others.

For second-order processes at the metacognitive level, moral conviction seems to inflate individuals' confidence in their judgments, regardless of whether those judgments were objectively correct. In fact, moral conviction increases confidence for incorrect judgments to a greater extent than it does for correct judgments. The mismatch between confidence and correctness was especially pronounced when the facts conflicted with individuals' strongly moralized beliefs, whereas no evidence was found that strong moral conviction enhances

metacognitive sensitivity for facts aligned with individuals' prior beliefs. These findings suggest that moral conviction not only motivates people to base factual judgments on views but also disrupts metacognitive processes, particularly when engaging with information that conflicts with people's beliefs. This metacognitive effect of moral conviction may further contribute to the increase in societal polarization. When moral conviction increases confidence in incorrect judgments, polarization may be sustained not only by motivated judgment itself, but also by people's heightened certainty that those judgments are correct. This may make individuals less willing to engage in cost-benefit analysis, update their beliefs, or consider compromise (O'Leary & Fletcher, 2024; Yoder & Decety, 2022).

Despite that, the current results support the claim that moral conviction can hurt factual judgment accuracy and metacognitive reflection, as discussed in depth (see Decety et al., 2025), it is not to say moral conviction is purely negative. It can also be socially constructive (Radke et al., 2020; Van Zomeren et al., 2012). In a study by Van Zomeren and colleagues (2011), it was found that among the advantaged social group, moral conviction about social inequality increases collective action tendencies through group-based anger and group efficacy. It was also found that heterosexual allies support LGBTQ rights when doing so aligns with their core moral values about justice and civil rights (Russell, 2011). This is especially important because much work on social change has focused on disadvantaged groups, for whom protest may be motivated by direct harm and injustice. By contrast, the advantaged group, while not always, often benefits from the existing status quo. Without moral conviction against inequality, they may have little motivation to challenge or change the system. For the disadvantaged group, moral conviction would also help sustain the collective effort to resist unjust authority or social pressure (Skitka et al., 2015). Therefore, while results from this dissertation caution us that moral conviction may

hurt judgments and metacognition, it is acknowledged that moral conviction, in many cases, can motivate positive changes, and sometimes may be necessary for those changes. Future studies should examine more closely the conditions under which moral conviction produces detrimental effects and the conditions under which it promotes positive social change. An important question is whether it is possible, and if it is, how to preserve the powerful motivational force of moral conviction while minimizing its potential epistemic costs, such as overconfidence and reduced belief updating, as well as social costs, such as increased polarization and distancing of disagreeing others.

In addition to examining moral conviction, the extent to which people ground their attitudes in what is right and wrong, this dissertation also considers the role of moral misalignment, that is, the extent to which an individual moralizes an attitude that conflicts with their beliefs. Across both first- and second-order processes, moral misalignment appears to play a particularly important role. In decision-making, such misalignment is associated with greater caution and slower responses. In factual judgments, moral misalignment impairs accuracy more than moral alignment improves it. Similarly, when it comes to metacognition, individuals' ability to distinguish their correct from incorrect judgments is especially poor when facts conflict with their strongly moralized beliefs, whereas their metacognitive sensitivity is not correspondingly enhanced when the fact aligns with those moralized beliefs. In other words, individuals' metacognitive judgments are not substantially better when the facts align with their strong moral convictions than when facts align with their personal preferences, whereas when facts conflict with strong moral convictions, confidence tracks accuracy significantly worse than when facts conflict with personal preferences. These results are in line with previous research showing that people are particularly sensitive to information that disconfirms their beliefs (Edwards & Smith,

1996), and they extend this effect to sociopolitical and metacognition domains and further identify the specific role of moral conviction in decision-making, factual judgments, and metacognitive evaluations. These findings highlight an important consideration for future research that the effects of aligned moral convictions on cognitive processes and behaviors may not be symmetrical with the effects of conflict with moral convictions. This is especially relevant to the effort to understand how to preserve the positive motivational force of moral conviction while reducing its negative consequences.

A limitation that applies across the studies in this dissertation is that moral conviction is rated by participants instead of being experimentally manipulated. While the use of socially and politically relevant issues increases the ecological validity of the present studies, it also introduces possible third variable confounds. Across studies, one especially plausible covariate, attitude strength, has been controlled and consistent with previous findings. Results show that moral significance can explain unique variance in neural activity, attention, and metacognitive judgments beyond the effects of attitude strength. However, other possible confounds, such as identity relevance, emotional investment, and familiarity, are not systematically controlled. However, it is also important to note that these variables may be the components of moral conviction or overlap with the moral conviction conceptually. Thus, the challenge is not just to control these factors, but to better characterize how moral conviction relates to the broader series of motivational, affective, and identity-relevant processes.

Overall, a mechanism that may connect all these findings is that moral conviction increases an issue's subjective importance and emotional salience. It also biases people toward belief-consistent judgments and disturbs their metacognitive abilities to realize such bias. Such reduced metacognitive sensitivity is likely to be primarily driven by situations when information

at hand conflicts with existing moral convictions. At a broader level, the findings suggest that moral conviction may contribute to sociopolitical conflict not only because people deeply care about these issues, but also because moralized topics may be processed and evaluated in a different way than those that are less morally charged.

Appendix C

Items for the knowledge metacognition task for high moral conviction topics

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
Abortion	At what gestational age does a human embryo's heart first start beating, on average?	6 weeks (4 weeks after conception)	12 weeks (10 weeks after conception)	Abortion-restrictive
	A study examined women who sought abortions in 2008-2010 but were unable to get one and thus carried the baby to term. One week after (being denied an abortion, 59% of these women still wished that they had been able to obtain one. One year after giving birth to a child, what percentage of these women still wished that they could have had an abortion?	7%	20%	Abortion-restrictive
	What proportion of all abortions are performed after 21 weeks of gestation?	1%	21%	Abortion-permissive
	According to a five-year study examining the health and socioeconomic consequences of having an abortion in the US, 97.5% of women reported that it was the right decision one week after the abortion, and five years later, that number ____.	Increased to 99%	Decreased to 87%	Abortion-permissive
Gun control	In 2021, what proportion of murders in the United States involve firearms	81%	35%	Pro-gun-control

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	A study from Johns Hopkins University found _____ after the state of Missouri loosened background check requirements for purchasing a gun?	The number of gun homicides in the state increased by about 25%, while the number of non-gun homicides was unchanged	The total number of homicides in the state was unchanged	Pro-gun-control
	A classic study from the 1980s showed that ____ of criminals fear armed citizens more than they fear the police.	57%	35%	Anti-gun-control
	According to a 2016 study that reviewed 25 types of state gun control laws, what was the overall effect on firearm death rates?	Equally likely to increase or decrease death rates	Mostly decrease death rates	Anti-gun-control
Death penalty	What proportion of criminals on death row are actually innocent?	4%	10%	Pro-death-penalty
	What is the failure rate for death by lethal injection from 1890 to 2010?	7.12%	15.6%	Pro-death-penalty
	True or false: the death penalty has been shown to reduce crime rates in the United States by acting as a deterrent.	False	True	Anti- death-penalty
	How do the costs of life in prison compare to the costs of capital punishment, considering factors such as legal costs and time in prison?	Capital punishment is up to 10x more expensive than a lifetime sentence	A lifetime sentence is up to 10x more expensive than capital punishment	Anti- death-penalty

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
Physician-assisted dying	A study from Belgium found that in 2007, _____ of patients whose doctors sped up their death via assisted dying either had not provided or were not capable of providing explicit consent for this choice	32%	10%	Anti-physician-assisted-dying
	According to a comprehensive review on physician-assisted dying, _____	There was no evidence of a reduction in rates of other forms of suicide after assisted dying was legalized. Instead, studies suggest increases in overall self-initiated deaths	Legalization of physician-assisted dying is associated with a significant reduction in other forms of suicide, suggesting that when physician-assisted dying is available, people choose that option instead of other methods of ending one's own life.	Anti-physician-assisted-dying
	Which of these is a criterion for physician-assisted dying to be legal in the United States?	The patient has a terminal diagnosis with less than 6 months to live	The patient is experiencing and is likely to continue experiencing severe chronic pain	Pro-physician-assisted-dying

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	One concern about physician-assisted dying is the “slippery slope” argument, that legalizing assisted dying for limited cases will expand to broader or non-voluntary cases. Which of the following research findings is accurate?	A 20-year review of Belgian assisted dying cases (2002–2023) found no strong evidence for a slippery slope; most of the increase was explained by demographic changes such as population aging	A 20-year review of Belgian assisted dying cases (2002–2023) found evidence that a slippery slope was happening; even after accounting for demographic changes, euthanasia prevalence rose substantially	Pro-physician-assisted-dying
Homosexuality being a trait people are born with	In how many species, other than humans, has same-sex attraction been documented?	450	27	Homosexuality is more of an innate trait
	According to the most recent research, genes account for __% of the factors involved in same-sex attraction.	8-25%	27-41%	Homosexuality is less of an innate trait
Transgender people having the option to change their physical characteristics	In a study of U.S. military service members and their families, what percentage of transgender individuals (adolescents and adults) who started cross-sex hormone treatment continued hormone treatment for at least four years?	70%	90%	Anti-change
	What proportion of patients regret getting a gender-affirming mastectomy (“top surgery”)?	1%	4%	Pro-change
Policies that address	Across the United States, how much higher are incarceration rates for black vs. white Americans?	6x	2x	Pro-policies-addressing-racial-inequality

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
racial inequality	How does affirmative action in college admissions affect the mostly-white classmates of its direct beneficiaries?	They tend to have more positive attitudes towards racial minorities	They tend to have more negative/resentful attitudes towards racial minorities	Pro-policies-addressing-racial-inequality
	A study of students admitted to UC San Diego in the early 1990s found that the graduation rate was _____ % for students admitted under affirmative action policies, compared to an overall graduation rate of 74%	57%	70%	Anti-policies-addressing-racial-inequality
	Which of these statements about economic and racial inequality is correct?	The median net worth of African-American households increased 60% from 2019 to 2022	The median net worth of African-American households was \$10,000 in 2022	Anti-policies-addressing-racial-inequality
Stricter enforcement of laws against illegal immigration	What proportion of immigrants are estimated to have come to the United States illegally?	23%	12%	Pro-stricter enforcement of laws against illegal immigrants
	Among all undocumented immigrants ages 25-64, _____ have <i>not</i> completed high school, compared with 8% of U.S. born adults in the same age range	47%	27%	Pro-stricter enforcement of laws against illegal immigrants

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	According to a study examining Texas crime rate data from 2012-2018, which of these statements about illegal immigration and crime is true?	A 10% increase in illegal immigrants share of the population is associated with a 1% percent <i>decrease</i> in violent crime convictions	A 10% increase in illegal immigrants share of the population is associated with a 8% percent <i>increase</i> in violent crime convictions	Anti-stricter enforcement of laws against illegal immigrants
	Which of the following statements is true?	In the last decade, undocumented immigrants have generated a \$100 billion surplus in the Social Security program	Undocumented workers typically do not contribute to the Social Security program because their wages are paid off-the-books	Anti-stricter enforcement of laws against illegal immigrants
Police having free rein to shoot suspects when they feel threatened	From 2015 to March 2024, in cases of police shootings, approximately __% of the victims were reported to be armed.	83%	57%	Pro-police-shooting
	According to a study of shooting incidents in Minneapolis, police shootings are followed by ____ in crime rates in areas located 0.1 to 0.5 miles from the police shooting site.	a persistent increase	a significant decrease	Anti-police-shooting
Reducing the use of the criminal	From a 2018 report, prisoners in the United States make up about __% of the world prison population.	20%	10%	Pro-reduction

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
justice system in dealing with societal problems	Roughly ____ U.S. adults—have an incarceration, conviction, or arrest record.	77 million	107 million	Anti-reduction
Reducing wealth inequality	What percentage of total wealth in the United States is currently held by the top 1% of the population, according to Federal Reserve statistics through mid-2023?	26.5%	46.5%	Anti-reduction-of-wealth-inequality
	The 2017 Tax Cuts and Jobs Act, passed under President Trump, reduced taxes on businesses and employees. What effect did this have on typical workers in the following years?	The average worker experienced an annual gain of \$750 in real wages	The average worker experienced an annual loss of \$500 in real wages	Anti-reduction-of-wealth-inequality
	In 2022, what was the ratio of pay for CEOs vs. the average nonsupervisory worker at their company, among the top 350 publicly-traded companies in the United States?	344:1	34:1	Pro-reduction-of-wealth-inequality
	From 1979 to 2019, average worker productivity in the United States increased by 59%. In that same period of time, by how much did inflation-adjusted wages for the typical (median) worker increase?	16%	45%	Pro-reduction-of-wealth-inequality
Legislation to combat climate change	What percentage of scientists believe that climate change is caused by humans?	99%	71%	Pro-legislation
	The International Renewable Energy Agency (IRENA) estimates that ____ of the world's electricity can and should come from renewable energy by 2050.	90%	60%	Pro-legislation

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	According to WHO estimates, climate change will kill ____ people each year worldwide between 2030 to 2050	250,000	950,000	Anti-legislation
	Forests act as important carbon sinks that absorb CO ₂ from the atmosphere. From 2000 to 2021, the amount of forest cover in the United States	Increased by 18 million acres	Decreased by 14 million acres	Anti-legislation
Privately-run healthcare	A Commonwealth Fund analysis ranked which country better on a measure of “care process”, which combines assessments of preventive care, safe care, coordinated care, and engagement and patient preferences?	The United States, where health care is primarily privately-run	France, where health care is primarily government-run	Pro-privately-run-healthcare
	Which country has been recognized for its high rates of preventive services, such as flu vaccinations and breast cancer screenings?	The United States, where health care is primarily privately run	Canada, where health care is publicly funded and universally accessible	Pro-privately-run-healthcare
	Which country has a higher rate of preventable deaths, i.e., deaths that could have been prevented by timely and effective care or better public health measures?	The United States, where health care is primarily privately-run	The United Kingdom, where health care is primarily government-run	Anti-privately-run-healthcare
	Among high-income Western countries, what proportion guarantees universal health care access?	91%	62%	Anti-privately-run-healthcare

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
Benefits to human health from animal research outweighing harms to the animals	According to a study on 2 major German university medical centers, approximately ____ of animal studies were never published, resulting in the unnecessary use and often euthanasia of animals without contributing to scientific knowledge	33%-42%	63%-76%	Pro-animal-research
	Assume that scientists' best guess (before animal testing) is that a new drug has a 70% chance of being safe in humans. A recent analysis indicated that on average, if such a drug is found to be safe for mice, what is the likelihood of it being safe for humans?	76%	95%	Anti-animal-research
Drug legalization and decriminalization	What percentage of drug-related arrests in the United States in 2022 were for possession offenses?	88%	44%	Pro-drug-legalization-and-decriminalization
	The city of Vancouver piloted a program in which they provided supervised injection sites to certain neighborhoods, where individuals could go and use opioids. What was the effect of this program?	A 26% decrease in overdose deaths surrounding injection sites	A 12% increase in overdose deaths surrounding injection sites	Pro-drug-legalization-and-decriminalization
	In 2020, Oregon decriminalized the possession of small amounts of illicit drugs. Since then, opioid-related deaths have:	Increased compared to the national average	Decreased compared to the national average	Anti-drug-legalization-and-decriminalization

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	Studies that examine car accident rates, comparing U.S. states in which marijuana is legal to states in which marijuana use is illegal, have most often found which outcome?	States where recreational marijuana is legal show an increased rate of traffic accidents	No effect on the rate of traffic accidents	Anti-drug-legalization-and-decriminalization
Separation of church and state	Studies show that student prayer in public schools is:	Associated with an increase in academic performance	Associated with a decline in academic performance	Anti-separation-of-church-and-state
	When was the phrase “In God We Trust” added to all US currency?	1955 because during the Cold War the US government wanted to emphasize America’s religiosity in contrast to the Soviet Union.	1789 because the founders of the United States wanted to emphasize the important role religion played among the American people	Pro-separation-of-church-and-state
Censorship of extremist speech	When Parler, a fringe right-wing social media network, was taken offline by large tech companies in 2021, what effect did it have on total user engagement across all similar fringe right-wing social media platforms?	The total use of fringe right-wing social media platforms remained unchanged	The total use of fringe right-wing social media platforms declined	Anti-censorship-of-extremist-speech
	In 2019, a popular community of Trump supporters (r/TheDonald) was banned from the Reddit platform. The community then migrated to a separate Web site. A textual analysis showed what type of change in the typical content of posts after this forced migration?	Increased levels of toxic and hateful content	Decreased levels of toxic and hateful content	Anti-censorship-of-extremist-speech

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	According to a Meta internal study published in 2023, what was the effect of banning popular spreaders of hate speech from the Facebook platform?	A broad decline in hate speech over the longer term, despite an initial backlash among those who were most closely connected	A persistent increase in hate speech among those most closely connected with the banned users, due to anger about censorship	Pro-censorship-of-extremist-speech
	What was the effect of a 2017 German law that required social media platforms to remove harmful Internet content, comparing far-right Twitter users from Germany vs. Austria where no such law was in place?	An 11% reduction in hateful posts in Germany with no evidence that German users were more likely to leave the platform	A 7% greater likelihood of German users leaving the platform, with little change in hateful content before vs. after the law took effect among those who stayed	Pro-censorship-of-extremist-speech
Equal pay for men and women	According to an analysis of data collected from 2009 to 2011, what percentage of the gender pay gap persists within specific careers when controlling for education and experience?	68%	32%	Pro-equal-pay
	A 2023 study examining data from Poland showed that female graduates in STEM fields (science, technology, engineering, and mathematics) earn _____ than male counterparts within the first year post graduation	Over 20% less	About 2% less	Pro-equal-pay

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	A 2016 report found what about the differences between men and women in leadership roles?	Men tend to be more inquisitive	There were no differences between male and female leaders on any measurable characteristic	Anti-equal-pay
	According to a 2025 report, nearly 80% of the U.S. gender pay gap is explained by _____	Women working fewer hours, taking more career breaks, and shifting into lower-paying occupations	Weak enforcement of equal pay laws and the absence of strong workplace protections	Anti-equal-pay
Women in the workforce	According to research, a child whose mother worked an average of 15 hours per week while the child was under age 16 would have a GPA ____ than a similar child whose mother did not work at all.	6.2 percent higher	6.2 percent lower	Pro women in the workforce
	Research shows that children whose mothers started working more during their preschool years were ____ likely to be overweight or obese between ages 7 and 11, compared to children whose mothers did not work.	more than 2.6 times	half as	Anti women in the workforce
Movements to reduce sexual harassment and assault	What percentage of sexual assault reports filed with police are typically shown to be false or exaggerated upon investigation?	7%	24%	Pro movements
	What percentage of female employees report that they have been sexually harassed by a colleague?	34%	64%	Anti movements
Legal protection for	In the past decade (2013-2023), the number of American minors employed in violation of child labor laws rose by:	415%	41.5%	Pro legal protection for workers' rights

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
workers' rights	Right-to-work laws make union formation much more difficult, and lead to reduced union participation. Compared to states without these laws, employees in right-to-work states report:	Higher levels of well-being on a "ladder" of life satisfaction	Lower levels of well-being on a "ladder" of life satisfaction	Anti legal protection for workers' rights
Using deadly force in self-defense when someone believes they are in imminent danger	According to a study published in 2001, robbery victims generally have a 30.2% chance of being injured; what percentage of victims who used a firearm for self-defense during a robbery were injured?	12.8%	41.6%	Pro deadly force
	What is the effect of the stand-your-ground laws on rates of homicide in a state?	8%-11% increase in homicides	3-5% decrease in homicides	Anti deadly force
Torture and enhanced interrogation techniques	Were "enhanced interrogation" techniques, used against terrorism suspects after the 9/11 attacks, found to be an effective method of obtaining accurate information?	No	Yes	Anti torture and enhanced interrogation techniques
	Which of the following was used as an "enhanced interrogation" technique in the post-9/11 era?	Being kept in a coffin-sized box for more than 24 hours	Being deprived of food for more than 24 hours	Anti torture and enhanced interrogation techniques
	Investigations into CIA detainment sites found that a total of ____ terrorism suspects were subjected to waterboarding, a form of torture, since 2001.	3	10	Pro torture and enhanced interrogation techniques
	A 2014 Senate Intelligence Committee report indicated that at least _____ detainees suspected of terrorism have been subject to "enhanced interrogation techniques" (waterboarding, beating, sleep deprivation) by the CIA from 2002 to 2009.	39	119	Pro torture and enhanced interrogation techniques

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
The US's military interventions in other parts of the world	A report examining US drone strikes in the Middle East and North Africa from 2009-2016 identified what percentage of those killed as civilians?	7% - 15%	21% - 28%	Pro US's military interventions in other parts of the world
	Which is a more accurate estimate of the number of Iraqi civilians killed in war between 2003 and 2023, following the U.S. invasion of Iraq?	186,000 - 210,000	27,500- 34,000	Anti US's military interventions in other parts of the world
Reducing the US's military spending	What is the largest expense of the US federal government?	The Social Security program	The Department of Defense (DoD) budget	Anti-reducing the US's military spending
	The US military procures much of its equipment from defense contractors, with which the government negotiates prices. What is the typical profit margin for defense contractors in recent years?	40%	10%	Pro-reducing the US's military spending
Welfare and social support programs	What percentage of recipients of government assistance in the U.S. are able-bodied adults without dependents?	8%	3%	Anti welfare and social support programs
	What is the rate of welfare fraud (applying for and accepting benefits without the requisite criteria) in the United States?	13.3%	4.7%	Anti welfare and social support programs
	Approximately how many people were lifted out of poverty in 2022 due to government benefits, according to the U.S. Census department data?	37,500,000	3,550,000	Pro welfare and social support programs
	Which United States government assistance program serves a larger population?	SNAP – benefits valid only for groceries	TANF – cash assistance for families with children	Pro welfare and social support programs

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
increasing large corporations' social responsibility and ethics	As of January 2024, what proportion of post-COVID inflation could be attributed to price increases that primarily enhance corporate profits rather than covering increased production costs?	35%	10%	Pro increasing large corporations' social responsibility and ethics
	Over the past fifteen years, what was the average rate of corporate donations to charity (as a proportion of their profits)?	0.9%	2.8%	Pro increasing large corporations' social responsibility and ethics
	What proportion of companies in a recent worldwide survey reported having a corporate ethics code?	80%	50%	Anti increasing large corporations' social responsibility and ethics
	Companies can voluntarily publish Environmental, Social, and Governance (ESG) disclosures, which measure factors like social impact, environmental impact, and work culture. The proportion of companies listed in the S&P 500 (i.e., the 500 most significant American companies) who publish ESGs in recent years has:	Risen significantly to 86%	Dropped significantly to 47%	Anti increasing large corporations' social responsibility and ethics
Gender-affirming care for minors (GACM)	According to research examining more than 3000 transgender and nonbinary youth (aged 13-17), those who wanted gender-affirming hormone treatment (GAHT) and received it showed _____ compared to those who wanted GAHT but were not able to receive it.	Lower rates of recent depression and less attempted suicide	Higher rates of recent depression and more attempted suicide	Pro-GACM

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	After receiving gender-affirming hormones, transgender youth reported a significant _____ in levels of general well-being (subjective sense of health, happiness and life satisfaction)	Increase	Decrease	Pro-GACM
	In transgender adolescents, what physical effects have studies found after one year of GnRH agonist therapy (a form of gender-affirming care that pauses puberty by suppressing sex hormone production)?	A significant increase in body fat percentage and BMI and a decrease in lean body mass percentage, a pattern associated with long-term health problems including increased risks for cardiovascular disease and diabetes.	No difference in physical characteristics including body fat percentage, BMI, and lean body mass percentage, suggesting that this treatment is unlikely to affect risk for cardiovascular disease or diabetes	Anti-GACM
	Which of the following statements is true:	Research evidence on long-term benefits of hormone treatment for neurocognitive development and psychosexual development in transgender youth under 18 remains unclear and limited	Research has established long-term benefits of hormone treatment on neurocognitive development and psychosexual development in transgender youth under 18	Anti-GACM
The need to reduce worldwide poverty	According to estimates by Pew Research Center, ___ in 10 people globally live on less than 10 dollars a day.	7	4	Pro-reducing worldwide poverty

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	The World Bank estimates that the proportion of people living in extreme poverty (defined as having less than \$2.15 a day, adjusting for inflation and differences in cost of living between countries) has decreased from 38% of the world's population in 1990 to what percentage of the world population in 2022?	9%	25%	Anti-reducing worldwide poverty
Criminal behavior is inherent (genetic) rather than environmental (influenced by surroundings and upbringing)	A study of adult male prisoners in New York State found what percentage had been subjected to severe abuse prior to age 12?	35%	23%	Anti-criminal behavior is inherent
	A recent analysis of crime data from a mid-1970s Universal Basic Income pilot program in Canada found that when all residents of a specific town were given unconditional cash payments, what happened to the crime rate, relative to other nearby towns?	A significant drop in both violent and property crime	No substantial change in crime rates for either violent crimes or property crimes	Anti-criminal behavior is inherent
	A large meta-analysis combining 51 studies of twins and adoptees, aimed at quantifying genetic vs. environmental determinants of antisocial behavior and criminality, estimated that what percentage of variability in these outcomes were due to genetics?	41%	20%	Pro-criminal behavior is inherent
	Compared to the general male population, how much more likely are adult male prisoners to meet the diagnostic criteria for psychopathy, a condition with a strong genetic basis?	15-25x	5-10x	Pro-criminal behavior is inherent

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
Homeless people are to blame for their own homelessness	A recent study found that ___ percent of adults sleeping in homeless shelters were employed during the same year they had experienced homelessness.	53%	17%	Anti-blaming homeless people for their own homelessness
	Studies have consistently shown that which approach is more effective for reducing long-term homelessness, and for reducing the likelihood of a previously-homeless person being hospitalized for physical or mental health issues?	Housing First approaches, where housing support is provided first without conditions, and then people are given support to address psychiatric and drug issues	Treatment First approaches, where psychiatric and drug issues must be addressed first before housing is made available	Anti-blaming homeless people for their own homelessness
	According to a government report from 2023, in San Francisco, what proportion of homeless people living in encampments declined offers of shelter or reported already having shelter or housing?	64%	32%	Pro-blaming homeless people for their own homelessness
	A 2008 survey asking US city governments to list the top three causes of homelessness in their cities found that the most common response (68% of surveyed cities) was which of the following?	Addiction	Low wages	Pro-blaming homeless people for their own homelessness
Doing more to end child labor	According to the International Labor Organization, 168 million children, almost 1 in 10 worldwide, are engaged in child labor. About ___ of them are working in hazardous conditions, such as working outside in extreme heat or exposed to toxic chemicals.	50%	30%	Pro-doing more to end child labor
	What is the minimum age to legally work in a coal mine in the United States?	18	14	Anti-doing more to end child labor

Appendix D

Items for the knowledge metacognition task for low moral conviction topics

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
Pineapple on pizza	Acidity and fattiness are two distinct flavor profiles, which are often paired in foods to balance them both. Which of the following is an example of such a pairing?	Pineapples and pizza	Salami and cheese	Pro-pineapple-on-pizza
	Which of these pizza toppings is more popular among Americans?	Mushrooms	Pineapple	Anti-pineapple-on-pizza
Adding sugar and cream to black coffee	True or false: while cream and sugar add flavor and texture to coffee, they reduce the anti-inflammatory and antioxidant properties that are good for health in coffee.	False	True	Pro sugar and cream in black coffee
	Which of the following statements about one's caloric intake from coffee is true?	One packet of sugar and a tablespoon of half and half contain more than 10x the calories of a cup of coffee	Adding sugar and cream to coffee is less healthy, but black coffee makes up about 85% of the calories in the cup	Anti sugar and cream in black coffee
Drinking decaf coffee	Consumption of which type of coffee is associated with higher levels of tiredness and headaches?	Decaffeinated coffee	Regular (caffeinated) coffee	Anti-decaf-coffee
	True or false: Caffeine withdrawal symptoms, including headache, fatigue, low mood, and irritability, were reduced by drinking decaf coffee—regardless of whether people knew it was decaf	True	False	Pro-decaf-coffee

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
Formal dress codes in the workplace	Studies show that, when compared to formal dress codes, casual dress codes:	Improve morale by creating a relaxed atmosphere in the workplace	Hurt morale by weakening the sense of professionalism and company pride in the workplace	Anti-formal-dress-codes-in-workplace
	According to a study, how do university students perceive instructors who dress more formally?	As having higher competence but lower warmth, with the reduced warmth decreasing students' intention to enroll in their class.	As having higher competence and equivalent warmth, with the increased competence encouraging students to enroll in their class.	Anti-formal-dress-codes-in-workplace
	In companies which switched from a formal dress code to a casual one, what was the impact?	Companies noted a significant increase in tardiness and flirtatious behavior	Companies noted a significant improvement in employer-employee relations and transparency	Pro- formal-dress-codes-in-workplace
	A recent study showed that _____ clothing was generally rated as more appropriate for professionals to wear at work, which led to them being perceived as more ethical in their jobs.	Business casual	Casual	Pro- formal-dress-codes-in-workplace
Using emojis	___% of employees think that emojis improve workplace communication	15%	5%	Pro-emojis

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	According to research from customer service platform Intercom, customers are:	4 times more likely to respond to an email from a business if the message contains emojis	Half as likely to respond to an email from a business if the message contains emojis	Pro-emojis
	A 2023 study found that in a corporate setting, a leader's use of positive emojis:	Enhances employees' creativity because they feel more respected and accepted	Decreases employees' creativity because positive emoji use makes the work environment feel inauthentic	Pro-emojis
	Gen Z individuals are more likely to think that:	Text messages are better than emojis for conveying intense feelings	Emojis are better than text messages for conveying intense feelings	Anti-emojis
	True or False: A study found that the receiver's ratings of the feelings being communicated by a specific negative emoji were significantly different from ratings given by the sender	True	False	Anti-emojis
	True or False: A study found that the receiver's ratings of the feelings being communicated by a specific negative emoji were significantly different from ratings given by the sender	True	False	Anti-emojis

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
Multi-tasking vs. single-tasking	A large study found that among children and young adults aged 7 to 29 years old, those who more often used different forms of technology simultaneously in their everyday lives (i.e., multitasking) showed _____ performance on a cognitive test that measured ability to multitask successfully?	Better	Worse	Pro-multitasking
	According to a study of junior high school students, higher levels of media multitasking were found to predict _____ ability to think creatively later on, measured by how many novel uses they could come up with for an everyday object	Lower	Higher	Anti-multitasking
Reading vs watching	Which activity leads to stronger and more widespread brain activity, promoting greater imagination and creativity?	Reading, because it requires the brain to translate written language from visual input to semantic meaning	Watching films, because it involves decoding both images and audio at the same time	Pro-reading
	Which leads to higher levels of memory for news content?	Reading a transcript of a broadcast media news story	Watching a television news segment	Pro-reading
	True or false: A study on language acquisition found that seventh and eighth graders who watched a closed-caption TV science series twice a week for 12 weeks demonstrated better word knowledge and information recall compared to those who learned the material by reading the textbook	True	False	Anti-reading

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	A study on language acquisition found that seventh and eighth graders who watched a closed-caption TV science series twice a week for 12 weeks demonstrated better word knowledge and information recall compared to those who learned the material by reading the textbook.	True	False	Anti-reading
Eating spicy food	Capsaicin (the active ingredient in spicy foods) has been linked to _____	Higher rates of gastrointestinal symptoms such as nausea and vomiting	Thickening of the stomach lining, which would suggest that it improves the stomach's ability to handle acids, enzymes, and microorganisms	Anti-spicy-food
	Which of these statements about the evolution of spicy food is true?	Plants evolved to produce capsaicin because the pain prevented mammals, including humans, from consuming them	Humans evolved to enjoy spiciness due to its antimicrobial properties	Anti-spicy-food
	What are the names of the chemicals released by the body in response to consuming spicy foods, and what feeling does it induce in the body/brain?	endorphins/dopamine, for a feeling of euphoria	Cortisol/adrenaline, for a feeling of stress	Pro-spicy-food

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	Which is true?	Spicy foods help with weight loss because capsaicin increases metabolism, which helps burn calories faster	Spicy foods make it harder to lose weight because they do not make you feel as full, making them easier to consume in excess	Pro-spicy-food
A cold shower is better for health than a hot shower	What temperature shower is better for boosting blood circulation and reducing inflammation?	Cold shower	Hot shower	Pro-cold shower
	___ showers are better for sleep, by relaxing muscles and lowering core temperature	Hot	Cold	Anti-cold shower
E-books vs. physical books	How much better is reading comprehension with physical books versus e-books?	8x	2-3x	Pro-physical books
	___ books are considered more accessible to people with visual impairment.	Electronic	Physical	Anti-physical books
IOS vs. Android	What is the most popular mobile operating system by market share worldwide?	Androids, at 70%	iPhones, at 55%	Pro-Android
	Which mobile operating system is known for its enhanced consumer privacy protections?	iPhone	Android	Anti-Android
Shoes on or off in the house	What proportion of dirt and dust in a home is carried in by shoes?	33%	47%	Anti-shoes off
	According to a recent University of Arizona study, ___ of shoe bottoms carry <i>E. Coli</i> , a toxic bacteria.	25%	10%	Pro-shoes off

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
Notes taken by hands or typed	How much faster is typing than writing while taking notes?	2x	They are about the same	Pro-typing
	What kind of note-taking results in the best memory retention?	Handwriting on paper	Typing on a keyboard	Anti-typing
Film vs digital for capturing images and videos	What kind of camera generally captures higher resolution images?	Film	Digital	Pro-film
	In 2017, what percentage of the top 100 movies in box office revenue were shot at least partly via digital cameras, as opposed to on film?	92%	67%	Anti-film
Digital music vs. vinyl records	Which music format has a wider dynamic range, reproducing more auditory detail?	Digital streaming	Vinyl records	Pro-digital
	Which music format uses a frequency range most similar to human hearing?	Vinyl records	Digital streaming	Anti-digital
Drinking room temperature vs. chilled water	Does drinking iced water or room temperature water help burn a few more calories and how?	Iced water helps burn calories because the body needs to spend extra energy maintaining its internal temperature after drinking iced water.	Iced water causes fewer calories to be burned at rest than room temperature water because cold temperature slows down baseline metabolism	Pro-chilled water
	True or false: people show poorer endurance in riding a stationary bike after drinking cold water vs. room temperature water according to sports research	False	True	Pro-chilled water

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	True or false: Drinking chilled water can aggravate migraines in those already prone to them	True	False	Anti-chilled water
	Which type of water is better to drink when you have a cold or flu?	Room temperature water	Ice water	Anti-chilled water
Listening to classical music has many cognitive and health benefits	True or false: A 2023 study found that one session listening to classical music for 50 minutes affects gene expression in ways that compensate for dysfunction in patients with mild cognitive impairment and Alzheimer's	True	False	Pro-classical music
	Which type of listening material leads to better performance on cognitive tasks involving declarative memory and processing speed in older adults?	Listening to upbeat classical music	Listening to white noise only	Pro-classical music
	Which is better for cognitive development in young children?	Singing along to a nursery rhyme	Listening to a piece of classical music	Anti-classical music
	A single exposure to which type of music can increase empathy for peers who struggle with mental health?	A pop song with anxiety themed lyrics	An upbeat classical music piece	Anti-classical music
Coffee vs tea	Which of these drinks tends to soothe rather than worsen an upset stomach?	Tea	Coffee	Anti-coffee
	Regularly consuming ____ is associated with significantly lower rates of cancer and cardiovascular mortality, especially in men	Tea	Coffee	Anti-coffee
	Which of the following provides faster mental alertness?	Coffee	Tea	Pro-coffee

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	Which is true:	The consumption of 2.5 cups of coffee/day minimizes the risk of Alzheimer's disease	The consumption of 2.5 cups of tea/day minimizes the risk of Alzheimer's disease	Pro-coffee
Subtitles when watching TV shows	How do closed captions in movies or TV shows affect memory and comprehension?	They improve memory and comprehension of a video, by reinforcing heard dialogue with written dialogue	They reduce memory and comprehension of a video, by taking attention away from the video and directing it towards text instead	Pro-subtitles
	A study found that Spanish students of English as a foreign language significantly improved their vocabulary after watching 1 hour of a TV drama under which conditions?	with English subtitles	with no subtitles	Pro-subtitles
	According to a survey in 2023, when watching TV in a language they know, ____ of US adults generally prefer to have subtitles on?	38%	58%	Anti-subtitles

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	Which is true about people learning English as a foreign language (EFL) who are dependent on subtitles when watching television shows?	EFL learners tend to incorporate incorrect information from erroneous subtitles into their understanding of the show they are watching	Erroneous subtitles have no effect on the ability of EFL learners to understand a show correctly	Anti-subtitles
Online dating	What percentage of Americans have married or been in a committed relationship with someone they met through online dating?	12%	27%	Anti-online dating
	True or false: When compared to couples that met via more traditional channels, marriages that start with online dating tend to be less satisfying and less stable	True	False	Anti-online dating
	A common perception is that dating apps are used for casual encounters rather than serious relationships. Studies show that ___% of users are <i>not</i> seeking casual encounters	70%	45%	Pro-online dating
	Studies have explored the different motives for individuals who download Tinder. Which of these motivations was associated with greater use of the app?	Seeking a relationship	Seeking sexual experience	Pro-online dating

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
Playing video games is beneficial	A study of 2,000 children who played video games for three or more hours per day found that:	The children who played video games had <i>more</i> problems with attention and <i>more</i> symptoms of depression compared to their peers, as reported by their parents	The children who played video games had <i>fewer</i> problems with attention and <i>fewer</i> symptoms of depression compared to their peers, as reported by their parents	Anti-video games
	True or false: playing video games is linked to more aggressive, violent behaviors in adolescents	True	False	Anti-video games
	After 15 hours of supervised video game playing over 5 weeks, adults aged 65 or older showed ____ compared to those who did not play	Enhanced cognitive functioning	No difference in cognitive functioning	Pro-video games
	Which of the following statements is true?	A recent analysis found that playing shooter video games improves spatial skills as much as formal (high school and university-level) courses aimed at enhancing these same skills do	A recent analysis found that non-gamers use less of their brain for pattern detection tasks, suggesting that gaming may impair brain efficiency in attention-demanding tasks	Pro-video games

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
Texting vs. calling	A 2012 study found that 7-to-12 year old girls receiving a soothing message from their mother via _____ showed reductions in cortisol (a stress hormone) and increases in oxytocin (a bonding hormone), while receiving a similar message via _____ did not produce these benefits	Phone call; Text-based instant message	Text-based instant message; Phone call	Calling over texting
	A study from 2021 found that interactions through ____ created stronger social bonds and no increase in awkwardness, compared to interactions through _____	Calling; texting	Texting; calling	Calling over texting
	For customer support, _____	70% of mobile users prefer texting over calling	55% of mobile users prefer calling over texting	Texting over calling
	A 2021 study showed which of the following about couples in long-distance relationships?	More frequent and responsive texting predicted significantly greater relationship satisfaction	More frequent voice calls predicted significantly greater relationship satisfaction	Texting over calling
Keeping pets benefits their caregivers	What effect does removal of pets from homes have on patients with pet allergic asthma?	It reduces asthmatic airway sensitivity more than optimal medications alone	It has no effect on asthmatic airway sensitivity relative to optimal medications alone	Anti-keeping pets

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	How do rates of depression compare between pet owners and non-owners?	Pet ownership is associated with twice the likelihood of having depression	Pet ownership is associated with a reduced likelihood of having depression	Anti-keeping pets
	According to a 2021 study on childhood development, children who grow up with a dog show	Significantly decreased risk of developmental delay in communication skills	Significantly increased risk of developmental delay in problem-solving skills	Pro-keeping pets
	According to a study of families in which a child has been diagnosed with autism, what is the typical effect of getting a dog?	Relieving the burden on primary carers (parents, guardians, etc.), by providing a secondary companion for children	Increasing the burden on primary carers (parents, guardians, etc.) by adding an additional dependent to take care of	Pro-keeping pets
Remote vs. in-person work	Research shows that remote work is, on average, __% less productive than in-person work	10-20%	5-10%	Pro-in-person
	A study examining the effect of remote work in Germany during the COVID pandemic found that employees who spent a higher proportion of their work hours working from home had	Lower job satisfaction	Higher job satisfaction	Pro-in-person
	According to research from the firm Boston Consulting Group, companies offering remote work have seen revenue growth at __ the rate of companies which did not offer remote work over the past three years	4x	half	Anti-in-person

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	_____ of employees feel that flexible work arrangements, such as remote work, increase employee morale	90%	40%	Anti-in-person
Summer vs. winter	A study from Spain found that the proportion of patients admitted to hospitals with respiratory illnesses who end up dying is higher in _____	Summer, because hot temperatures aggravate chronic and infectious respiratory diseases	Winter, because cold temperatures aggravate chronic and infectious respiratory diseases	Winter over summer
	_____ can train blood vessels to be more responsive, leading to more vibrant, rosy skin over time	Moderately cold weather	Moderately hot weather	Winter over summer
	Summers are marked by longer days and shorter nights, when compared to winter. How does this impact sleep quality? Info: Melatonin is the primary hormone which causes sleepiness	Sleep quality is better in the summer, because increased vitamin D exposure during the day increases melatonin levels during the night ¹	Sleep quality is worse in the summer, because increased vitamin D exposure during the day decreases melatonin levels during the night	Summer over winter

¹ This answer is erroneous. Research has shown that longer summer days typically decreases or delay melatonin production. This item has been excluded from the analyses.

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	According to a review paper that analyzed studies on the relationship between seasonality and physical activity, _____	17 out of 26 studies found that physical activity levels are generally higher in summer than in winter, while the remaining studies did not find significant effects	15 out of 26 studies found no significant relationship between season and physical activity levels, while the remaining studies showed mixed results (i.e., some in either direction)	Summer over winter
Living in an urban vs. rural area	True or false: individuals living in urban environments are at a higher risk of psychiatric disorder than individuals living in rural environments.	False	True	Pro-urban
	True or false: in less developed countries where GDP per capita is below \$20,000, rates of wellbeing are higher for urban populations than for rural populations	True	False	Pro-urban
	What impact does urbanization have on public health, according to a study using annual panel data published by the World Bank from 2000 to 2018?	Positive impact, because urbanization increases living standards and overall life expectancy	Negative impact, because rapid urbanization decreases living standards and increases obesity rate	Pro-urban

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	Which environment is associated with a higher sense of community belonging?	Rural communities, due to tight-knit, small social circles	Urban communities, due to increased options for socializing and geographic proximity to friends and family	Anti-urban
	Which of the following statements is true?	Compared to rural residents, urban residents are more likely to want to move to a different community type	Compared to urban residents, rural residents are more likely to want to move to a different community type	Anti-urban
	Which of the following improves executive attention in both older and younger adults?	Viewing images of nature	Viewing images of urban environments	Anti-urban
Driving vs. public transportation	When compared to driving, taking public transportation decreases household expenditure by ___ \$ annually	\$13,000	\$4,000	Pro-public transport
	Which of the following statements is true?	The rates of accident-related deaths are 60x lower when commuting by bus than when commuting by car	Passengers on public transit are 3x more likely to be the victim of a crime as compared to car drivers	Pro-public transport

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	A study found that users describe which type of transportation as providing higher levels of physical and social protection and affirmation of self-worth?	Cars	Public transport	Anti-public transport
	In the United States, on average, public transport commute times are:	On average twice as long as driving times	On average 1.5 times as long as driving times	Anti-public transport
Night owls vs. early birds	Who is at greater risk of developing diabetes and other health problems, even after controlling for lifestyle factors?	People with later sleep and wake times	People with early sleep and wake times	Early birds over night owls
	What percentage of successful CEOs and businesspeople get up before 6 am every day, according to a survey by Inc. magazine?	69%	52%	Early birds over night owls
	A 2009 longitudinal study examining the relationship between childhood intelligence and adult circadian rhythms found that children with higher IQ scores tended, as adults, to go to sleep and wake up _____ than those with lower childhood IQ scores	Later	Earlier	Night owls over early birds
	A 2008 study found that 58% of art students in university feel more creative _____.	after 10pm	Before 10am	Night owls over early birds

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
Binge-watching TV series vs. watching them over time	Is binge watching better or worse than spaced viewing for recall 4 weeks later of episodes from a TV series?	Better, because binge watching benefits from hyper-attention and uninterrupted immersion into a narrative that allows for enriched encoding and better retrieval later.	Worse, because binge watching causes cognitive overload, giving the brain too much information in too little time to organize and store information for later recall.	Pro-binge-watching
	Which is better for learning the vocabulary of a second-language from TV?	Spaced viewing of TV show episodes	Binge watching of TV show episodes	Anti-binge-watching
Showering vs. bathing	In a study from Japan comparing routine warm-water immersion bathing with showering, where people were randomly assigned to 2 weeks of each, which one led to greater reductions in fatigue, less self-reported stress, and better mental health?	Bathing	Showering	Bathing over showering

Topic	Question	Correct Answer	Incorrect Answer	Position Favored by the Correct Answer
	Which of the following is true:	For people with chronic heart failure, taking warm baths can help improve heart function by relaxing blood vessels.	For people with chronic heart failure, taking warm showers has been shown in clinical trials to be more effective than warm-water bathing in improving heart function.	Bathing over showering
	While both showers and baths have their benefits, ____ have a clear advantage when it comes to getting cleaner.	Showers, by continually replacing dirty water with clean water as it washes off.	Baths, by immersing the body and ensuring near-total coverage as one cleans.	Showering over bathing
	How much more water does a typical bath use compared to a 5-minute shower?	3x	1.5x	Showering over bathing

REFERENCES

- Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, 191, 104254. <https://doi.org/10.1016/j.jpubeco.2020.104254>
- Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, 20(2), 870–888. [https://doi.org/10.1016/S1053-8119\(03\)00336-7](https://doi.org/10.1016/S1053-8119(03)00336-7)
- Andrejević, M., White, J. P., Feuerriegel, D., Laham, S., & Bode, S. (2022). Response time modelling reveals evidence for multiple, distinct sources of moral decision caution. *Cognition*, 223, 105026. <https://doi.org/10.1016/j.cognition.2022.105026>
- Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83(6), 1423–1440. <https://doi.org/10.1037/0022-3514.83.6.1423>
- Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26–41. <https://doi.org/10.1016/j.media.2007.06.004>
- Avramova, Y. R., & Inbar, Y. (2013). Emotion and moral judgment. *WIREs Cognitive Science*, 4(2), 169–178. <https://doi.org/10.1002/wcs.1216>
- Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, 76, 412–427. <https://doi.org/10.1016/j.neuroimage.2013.02.063>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is Stronger than Good. *Review of General Psychology*, 5(4), 323–370. <https://doi.org/10.1037/1089-2680.5.4.323>
- Baumert, A., Beierlein, C., Schmitt, M., Kemper, C. J., Kovaleva, A., Liebig, S., & Rammstedt, B. (2014). Measuring Four Perspectives of Justice Sensitivity With Two Items Each. *Journal of Personality Assessment*, 96(3), 380–390. <https://doi.org/10.1080/00223891.2013.836526>
- Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience*, 14(11), 1468–1474. <https://doi.org/10.1038/nn.2933>
- Behzadi, Y., Restom, K., Liao, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1), 90–101. <https://doi.org/10.1016/j.neuroimage.2007.04.042>

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Boxell, L., Gentzkow, M., & Shapiro, J. M. (2024). Cross-Country Trends in Affective Polarization. *Review of Economics and Statistics*, 106(2), 557–565. https://doi.org/10.1162/rest_a_01160
- Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The MAD Model of Moral Contagion: The Role of Motivation, Attention, and Design in the Spread of Moralized Content Online. *Perspectives on Psychological Science*, 15(4), 978–1010. <https://doi.org/10.1177/1745691620917336>
- Brady, W. J., Gantman, A. P., & Van Bavel, J. J. (2020). Attentional capture helps explain why moral and emotional content go viral. *Journal of Experimental Psychology: General*, 149(4), 746–756. <https://doi.org/10.1037/xge0000673>
- 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. <https://doi.org/10.1073/pnas.1618923114>
- Brandt, M. J., Wisneski, D. C., & Skitka, L. J. (2015). Moralization and the 2012 U.S. presidential election campaign. *Journal of Social and Political Psychology*, 3(2), 211–237. <https://doi.org/10.5964/jspp.v3i2.434>
- Calvo, M. G., & Lang, P. J. (2004). Gaze Patterns When Looking at Emotional Pictures: Motivationally Biased Attention. *Motivation and Emotion*, 28(3), 221–243. <https://doi.org/10.1023/B:MOEM.0000040153.26156.ed>
- Cao, Q., Cohen, M. S., Bakkour, A., Leong, Y. C., & Decety, J. (2025). Moral conviction interacts with metacognitive ability in modulating neural activity during sociopolitical decision-making. *Cognitive, Affective, & Behavioral Neuroscience*, 25(2), 291–310. <https://doi.org/10.3758/s13415-024-01243-3>
- Carcione, A., Riccardi, I., Bilotta, E., Leone, L., Pedone, R., Conti, L., Colle, L., Fiore, D., Nicolò, G., Pellicchia, G., Procacci, M., & Semerari, A. (2019). Metacognition as a Predictor of Improvements in Personality Disorders. *Frontiers in Psychology*, 10, 170. <https://doi.org/10.3389/fpsyg.2019.00170>
- Carlson, R. W., & Crockett, M. J. (2018). The lateral prefrontal cortex and moral goal pursuit. *Current Opinion in Psychology*, 24, 77–82. <https://doi.org/10.1016/j.copsyc.2018.09.007>
- Chen, T., Cai, W., Ryali, S., Supekar, K., & Menon, V. (2016). Distinct Global Brain Dynamics and Spatiotemporal Organization of the Salience Network. *PLOS Biology*, 14(6), e1002469. <https://doi.org/10.1371/journal.pbio.1002469>

- Ciuk, D. J., & Rottman, J. (2021). Moral Conviction, Emotion, and the Influence of Episodic versus Thematic Frames. *Political Communication*, 38(5), 519–538. <https://doi.org/10.1080/10584609.2020.1793847>
- Clifford, S. (2019). How Emotional Frames Moralize and Polarize Political Attitudes. *Political Psychology*, 40(1), 75–91. <https://doi.org/10.1111/pops.12507>
- Clithero, J. A., Smith, D. V., Carter, R. M., & Huettel, S. A. (2011). Within- and cross-participant classifiers reveal different neural coding of information. *NeuroImage*, 56(2), 699–708. <https://doi.org/10.1016/j.neuroimage.2010.03.057>
- Cousar, K. A., Carnes, N. C., & Kimel, S. Y. (2021). Morality as Fuel for Violence? Disentangling the Role of Religion in Violent Conflict. *Social Cognition*, 39(1), 166–182. <https://doi.org/10.1521/soco.2021.39.1.166>
- Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, 10(4–5), 171–178. [https://doi.org/10.1002/\(SICI\)1099-1492\(199706/08\)10:4/5%3C171::AID-NBM453%3E3.0.CO;2-L](https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5%3C171::AID-NBM453%3E3.0.CO;2-L)
- Crockett, M. J., Siegel, J. Z., Kurth-Nelson, Z., Dayan, P., & Dolan, R. J. (2017). Moral transgressions corrupt neural representations of value. *Nature Neuroscience*, 20(6), 879–885. <https://doi.org/10.1038/nn.4557>
- Cunningham, W. A., & Brosch, T. (2012). Motivational Salience: Amygdala Tuning From Traits, Needs, Values, and Goals. *Current Directions in Psychological Science*, 21(1), 54–59. <https://doi.org/10.1177/0963721411430832>
- Cusimano, C., & Lombrozo, T. (2021). Morality justifies motivated reasoning in the folk ethics of belief. *Cognition*, 209, 104513. <https://doi.org/10.1016/j.cognition.2020.104513>
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, 9(2), 179–194. <https://doi.org/10.1006/nimg.1998.0395>
- De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in Doubt: Conflict and Decision Confidence. *PLoS ONE*, 6(1), e15954. <https://doi.org/10.1371/journal.pone.0015954>
- Decety, J. (2024). *The power of moral conviction: How it catalyzes dogmatism, intolerance, and violence*. <https://doi.org/10.5281/ZENODO.11001101>
- Decety, J., Cohen, M. S., & Cao, Q. (2025). The Dark Side of Moral Conviction—Integrating Political Psychology, Cognitive Science, and Neuroscience. *Annals of the New York Academy of Sciences*, 1553(1), 5–20. <https://doi.org/10.1111/nyas.70109>
- Decety, J., & Cowell, J. M. (2018). Interpersonal harm aversion as a necessary foundation for morality: A developmental neuroscience perspective. *Development and Psychopathology*, 30(1), 153–164. <https://doi.org/10.1017/S0954579417000530>

- Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Chapter 10 Motivated Moral Reasoning. In *Psychology of Learning and Motivation* (Vol. 50, pp. 307–338). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)00410-6](https://doi.org/10.1016/S0079-7421(08)00410-6)
- Dixon, M. L., & Christoff, K. (2014). The lateral prefrontal cortex and complex value-based learning and decision making. *Neuroscience & Biobehavioral Reviews*, 45, 9–18. <https://doi.org/10.1016/j.neubiorev.2014.04.011>
- Dobbs, M., DeGutis, J., Morales, J., Joseph, K., & Swire-Thompson, B. (2023). Democrats are better than Republicans at discerning true and false news but do not have better metacognitive awareness. *Communications Psychology*, 1(1), 46. <https://doi.org/10.1038/s44271-023-00040-x>
- Dunlap, R. E., McCright, A. M., & Yarosh, J. H. (2016). The Political Divide on Climate Change: Partisan Polarization Widens in the U.S. *Environment: Science and Policy for Sustainable Development*, 58(5), 4–23. <https://doi.org/10.1080/00139157.2016.1208995>
- Duverno, S., & Koechlin, E. (2017). Rewards and Cognitive Control in the Human Prefrontal Cortex. *Cerebral Cortex*, 27(10), 5024–5039. <https://doi.org/10.1093/cercor/bhx210>
- Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, 71(1), 5–24. <https://doi.org/10.1037/0022-3514.71.1.5>
- Epley, N., & Gilovich, T. (2016). The Mechanics of Motivated Reasoning. *Journal of Economic Perspectives*, 30(3), 133–140. <https://doi.org/10.1257/jep.30.3.133>
- Eres, R., Louis, W. R., & Molenberghs, P. (2018). Common and distinct neural networks involved in fMRI studies investigating morality: An ALE meta-analysis. *Social Neuroscience*, 13(4), 384–398. <https://doi.org/10.1080/17470919.2017.1357657>
- Erisen, C., Redlawsk, D. P., & Erisen, E. (2018). Complex Thinking as a Result of Incongruent Information Exposure. *American Politics Research*, 46(2), 217–245. <https://doi.org/10.1177/1532673X17725864>
- Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLOS ONE*, 12(9), e0184661. <https://doi.org/10.1371/journal.pone.0184661>
- Esteban, O., Ciric, R., Finc, K., Blair, R. W., Markiewicz, C. J., Moodie, C. A., Kent, J. D., Goncalves, M., DuPre, E., Gomez, D. E. P., Ye, Z., Salo, T., Valabregue, R., Amlie, I. K., Liem, F., Jacoby, N., Stojić, H., Cieslak, M., Urchs, S., ... Gorgolewski, K. J. (2020). Analysis of task-based functional MRI data preprocessed with fMRIPrep. *Nature Protocols*, 15(7), 2186–2202. <https://doi.org/10.1038/s41596-020-0327-3>
- Evans, A. C., Janke, A. L., Collins, D. L., & Baillet, S. (2012). Brain templates and atlases. *NeuroImage*, 62(2), 911–922. <https://doi.org/10.1016/j.neuroimage.2012.01.024>

- Feinberg, M., Kovacheff, C., Teper, R., & Inbar, Y. (2019). Understanding the process of moralization: How eating meat becomes a moral issue. *Journal of Personality and Social Psychology*, 117(1), 50–72. <https://doi.org/10.1037/pspa0000149>
- FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., & Mobbs, D. (2012). Differential neural circuitry and self-interest in real vs hypothetical moral decisions. *Social Cognitive and Affective Neuroscience*, 7(7), 743–751. <https://doi.org/10.1093/scan/nss069>
- FeldmanHall, O., & Mobbs, D. (2015). A Neural Network for Moral Decision Making. In *Brain Mapping* (pp. 205–210). Elsevier. <https://doi.org/10.1016/B978-0-12-397025-1.00180-9>
- Fiedler, S., & Glöckner, A. (2015). Attention and moral behavior. *Current Opinion in Psychology*, 6, 139–144. <https://doi.org/10.1016/j.copsyc.2015.08.008>
- Finn, B., & Tauber, S. K. (2015). When Confidence Is Not a Signal of Knowing: How Students' Experiences and Beliefs About Processing Fluency Can Lead to Miscalibrated Confidence. *Educational Psychology Review*, 27(4), 567–586. <https://doi.org/10.1007/s10648-015-9313-7>
- Fischer, H., & Fleming, S. (2024). Why metacognition matters in politically contested domains. *Trends in Cognitive Sciences*, 28(9), 783–785. <https://doi.org/10.1016/j.tics.2024.06.005>
- Fischer, H., & Said, N. (2021). Importance of domain-specific metacognition for explaining beliefs about politicized science: The case of climate change. *Cognition*, 208, 104545. <https://doi.org/10.1016/j.cognition.2020.104545>
- Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, 38(6), 889–906. <https://doi.org/10.1037/0022-3514.38.6.889>
- Fleming, S. M. (2017). HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, 2017(1). <https://doi.org/10.1093/nc/nix007>
- Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1338–1349. <https://doi.org/10.1098/rstb.2011.0417>
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00443>
- Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, 137(10), 2811–2822. <https://doi.org/10.1093/brain/awu221>

- Forbes, C. E., & Grafman, J. (2010). The Role of the Human Prefrontal Cortex in Social Cognition and Moral Judgment. *Annual Review of Neuroscience*, 33(1), 299–324. <https://doi.org/10.1146/annurev-neuro-060909-153230>
- Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of Psychology*, 67(1), 641–666. <https://doi.org/10.1146/annurev-psych-122414-033645>
- Gantman, A., Devraj-Kizuk, S., Mende-Siedlecki, P., Van Bavel, J. J., & Mathewson, K. E. (2020). The time course of moral perception: An ERP investigation of the moral pop-out effect. *Social Cognitive and Affective Neuroscience*, 15(2), 235–246. <https://doi.org/10.1093/scan/nsaa030>
- Garrett, K. N. (2019). Fired Up by Morality: The Unique Physiological Response Tied to Moral Conviction in Politics. *Political Psychology*, 40(3), 543–563. <https://doi.org/10.1111/pops.12527>
- 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. <https://doi.org/10.1017/S000712341700059X>
- Gehring, W. J., & Willoughby, A. R. (2002). The Medial Frontal Cortex and the Rapid Processing of Monetary Gains and Losses. *Science*, 295(5563), 2279–2282. <https://doi.org/10.1126/science.1066893>
- Ghaffari, M., & Fiedler, S. (2018). The Power of Attention: Using Eye Gaze to Predict Other-Regarding and Moral Choices. *Psychological Science*, 29(11), 1878–1889. <https://doi.org/10.1177/0956797618799301>
- Golman, R. (2023). Acceptable discourse: Social norms of beliefs and opinions. *European Economic Review*, 160, 104588. <https://doi.org/10.1016/j.euroecorev.2023.104588>
- Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, 106(3), 1339–1366. <https://doi.org/10.1016/j.cognition.2007.06.007>
- Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences*, 106(30), 12506–12511. <https://doi.org/10.1073/pnas.0900152106>
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1), 63–72. <https://doi.org/10.1016/j.neuroimage.2009.06.060>
- Guggenmos, M. (2021). Measuring metacognitive performance: Type 1 performance dependence and test-retest reliability. *Neuroscience of Consciousness*, 2021(1), niab040. <https://doi.org/10.1093/nc/niab040>

- Hamilton, L. (2022). *Conspiracy vs. Science: A Survey of U.S. Public Beliefs*. Carsey School of Public Policy, University of New Hampshire.
<https://carsey.unh.edu/publication/conspiracy-vs-science-survey-us-public-beliefs>
- Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., & Rangel, A. (2010). Value Computations in Ventral Medial Prefrontal Cortex during Charitable Decision Making Incorporate Input from Regions Involved in Social Cognition. *The Journal of Neuroscience*, 30(2), 583–590. <https://doi.org/10.1523/JNEUROSCI.4089-09.2010>
- Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System. *Science*, 324(5927), 646–648.
<https://doi.org/10.1126/science.1168450>
- Hare, T. A., Hakimi, S., & Rangel, A. (2014). Activity in dlPFC and its effective connectivity to vmPFC are associated with temporal discounting. *Frontiers in Neuroscience*, 8.
<https://doi.org/10.3389/fnins.2014.00050>
- Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors. *The Journal of Neuroscience*, 28(22), 5623–5630.
<https://doi.org/10.1523/JNEUROSCI.1309-08.2008>
- Heinzelmann, N., Hölting, B. T. A., & Tran, V. (2021). Moral discourse boosts confidence in moral judgments. *Philosophical Psychology*, 34(8), 1192–1216.
<https://doi.org/10.1080/09515089.2021.1959026>
- Heinzelmann, N., & Tran, V. (2024). Extremists are more confident. *Erkenntnis*, 89(5), 2031–2056. <https://doi.org/10.1007/s10670-022-00616-9>
- Hesse, E., Mikulan, E., Decety, J., Sigman, M., Garcia, M. D. C., Silva, W., Ciraolo, C., Vaucheret, E., Baglivo, F., Huepe, D., Lopez, V., Manes, F., Bekinschtein, T. A., & Ibanez, A. (2016). Early detection of intentional harm in the human amygdala. *Brain*, 139(1), 54–61. <https://doi.org/10.1093/brain/awv336>
- Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F. S., & Behrens, T. E. J. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nature Neuroscience*, 15(3), 470–476. <https://doi.org/10.1038/nn.3017>
- Hutcherson, C. A., Montaser-Kouhsari, L., Woodward, J., & Rangel, A. (2015). Emotional and Utilitarian Appraisals of Moral Dilemmas Are Encoded in Separate Areas and Integrated in Ventromedial Prefrontal Cortex. *Journal of Neuroscience*, 35(36), 12593–12605.
<https://doi.org/10.1523/JNEUROSCI.3402-14.2015>
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, 17(2), 825–841. <https://doi.org/10.1006/nimg.2002.1132>

- Kaanders, P., Sepulveda, P., Folke, T., Ortoleva, P., & De Martino, B. (2022). Humans actively sample evidence to support prior beliefs. *eLife*, *11*, e71768. <https://doi.org/10.7554/eLife.71768>
- Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*(4), 407–424. <https://doi.org/10.1017/S1930297500005271>
- Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. <https://doi.org/10.1371/journal.pcbi.1005350>
- Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2020). Confirmation Bias, Ingroup Bias, and Negativity Bias in Selective Exposure to Political Information. *Communication Research*, *47*(1), 104–124. <https://doi.org/10.1177/0093650217719596>
- Knoch, D., Schneider, F., Schunk, D., Hohmann, M., & Fehr, E. (2009). Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proceedings of the National Academy of Sciences*, *106*(49), 20895–20899. <https://doi.org/10.1073/pnas.0911619106>
- Koenig, H. G., & Büssing, A. (2010). The Duke University Religion Index (DUREL): A Five-Item Measure for Use in Epidemiological Studies. *Religions*, *1*(1), 78–85. <https://doi.org/10.3390/rel1010078>
- 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. <https://doi.org/10.1111/sipr.12045>
- Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298. <https://doi.org/10.1038/nn.2635>
- Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions. *Frontiers in Psychology*, *3*. <https://doi.org/10.3389/fpsyg.2012.00193>
- Krueger, F., & Hoffman, M. (2016). The Emerging Neuroscience of Third-Party Punishment. *Trends in Neurosciences*, *39*(8), 499–501. <https://doi.org/10.1016/j.tins.2016.06.004>
- Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, *108*(3), 480–498.
- Laborde, C. (with Bardon, A.). (2017). *Religion in Liberal Political Philosophy*. Oxford University Press USA - OSO.
- Lanczos, C. (1964). Evaluation of Noisy Data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, *1*(1), 76–85. <https://doi.org/10.1137/0701007>

- Lapate, R. C., Samaha, J., Rokers, B., Postle, B. R., & Davidson, R. J. (2020). Perceptual metacognition of human faces is causally supported by function of the lateral prefrontal cortex. *Communications Biology*, 3(1), 360. <https://doi.org/10.1038/s42003-020-1049-3>
- Lee, N., Nyhan, B., Reifler, J., & Flynn, D. J. (2021). More Accurate, But No Less Polarized: Comparing the Factual Beliefs of Government Officials and the Public. *British Journal of Political Science*, 51(3), 1315–1322. <https://doi.org/10.1017/S000712342000037X>
- Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467–477. <https://doi.org/10.1121/1.1912375>
- Levy, D. J., & Glimcher, P. W. (2011). Comparing Apples and Oranges: Using Reward-Specific and Reward-General Subjective Value Representation in the Brain. *The Journal of Neuroscience*, 31(41), 14693–14707. <https://doi.org/10.1523/JNEUROSCI.2218-11.2011>
- Lin, H., Müller-Bardorff, M., Gathmann, B., Brieke, J., Mothes-Lasch, M., Bruchmann, M., Miltner, W. H. R., & Straube, T. (2020). Stimulus arousal drives amygdalar responses to emotional expressions across sensory modalities. *Scientific Reports*, 10(1), 1898. <https://doi.org/10.1038/s41598-020-58839-1>
- Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2018). The role of a “common is moral” heuristic in the stability and change of moral norms. *Journal of Experimental Psychology: General*, 147(2), 228–242. <https://doi.org/10.1037/xge0000365>
- Liu, B. S., & Ditto, P. H. (2013). What Dilemma? Moral Evaluation Shapes Factual Belief. *Social Psychological and Personality Science*, 4(3), 316–323. <https://doi.org/10.1177/1948550612456045>
- Lodge, M., & Taber, C. S. (2005). The Automaticity of Affect for Political Leaders, Groups, and Issues: An Experimental Test of the Hot Cognition Hypothesis. *Political Psychology*, 26(3), 455–482. <https://doi.org/10.1111/j.1467-9221.2005.00426.x>
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109. <https://doi.org/10.1037/0022-3514.37.11.2098>
- 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. <https://doi.org/10.1016/j.jesp.2016.04.003>
- Luttrell, A., & Togans, L. J. (2021). The Stability of Moralized Attitudes Over Time. *Personality and Social Psychology Bulletin*, 47(4), 551–564. <https://doi.org/10.1177/0146167220935737>
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, 21(1), 422–430. <https://doi.org/10.1016/j.concog.2011.09.021>

- Marcus, G. E., & MacKuen, M. B. (1993). Anxiety, Enthusiasm, and the Vote: The Emotional Underpinnings of Learning and Involvement During Presidential Campaigns. *American Political Science Review*, 87(3), 672–685. <https://doi.org/10.2307/2938743>
- Marie, A., Altay, S., & Strickland, B. (2023). Moralization and extremism robustly amplify myside sharing. *PNAS Nexus*, 2(4), pgad078. <https://doi.org/10.1093/pnasnexus/pgad078>
- Mazancieux, A., Fleming, S. M., Souchay, C., & Moulin, C. J. A. (2020). Is there a G factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks. *Journal of Experimental Psychology: General*, 149(9), 1788–1799. <https://doi.org/10.1037/xge0000746>
- Mazancieux, A., Pereira, M., Faivre, N., Mamassian, P., Moulin, C. J. A., & Souchay, C. (2023). Towards a common conceptual space for metacognition in perception and memory. *Nature Reviews Psychology*, 2(12), 751–766. <https://doi.org/10.1038/s44159-023-00245-1>
- McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *NeuroImage*, 61(4), 1277–1286. <https://doi.org/10.1016/j.neuroimage.2012.03.068>
- Meppelink, C. S., Smit, E. G., Fransen, M. L., & Diviani, N. (2019). “I was Right about Vaccination”: Confirmation Bias and Health Literacy in Online Health Information Seeking. *Journal of Health Communication*, 24(2), 129–140. <https://doi.org/10.1080/10810730.2019.1583701>
- Mesbah, H. M. (2005). Reading is Remembering: The Effect of Reading vs. Watching News on Memory and Metamemory. *Speaker & Gavel*, 42(1). <https://cornerstone.lib.mnsu.edu/speaker-gavel/vol42/iss1/6>
- Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4), 851–861. <https://doi.org/10.1037/0278-7393.19.4.851>
- Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, 24(1), 167–202. <https://doi.org/10.1146/annurev.neuro.24.1.167>
- Mormann, F., Bausch, M., Knieling, S., & Fried, I. (2019). Neurons in the Human Left Amygdala Automatically Encode Subjective Value Irrespective of Task. *Cerebral Cortex*, 29(1), 265–272. <https://doi.org/10.1093/cercor/bhx330>
- Mulckhuyse, M. (2018). The influence of emotional stimuli on the oculomotor system: A review of the literature. *Cognitive, Affective, & Behavioral Neuroscience*, 18(3), 411–425. <https://doi.org/10.3758/s13415-018-0590-8>

- Muschelli, J., Nebel, M. B., Caffo, B. S., Barber, A. D., Pekar, J. J., & Mostofsky, S. H. (2014). Reduction of motion-related artifacts in resting state fMRI using aCompCor. *NeuroImage*, 96, 22–35. <https://doi.org/10.1016/j.neuroimage.2014.03.028>
- Neuman, S. B., & Koskinen, P. (1992). Captioned Television as Comprehensible Input: Effects of Incidental Word Learning from Context for Language Minority Students. *Reading Research Quarterly*, 27(1), 94. <https://doi.org/10.2307/747835>
- Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nucci, L. P., & Turiel, E. (1978). Social Interactions and the Development of Social Concepts in Preschool Children. *Child Development*, 49(2), 400. <https://doi.org/10.2307/1128704>
- O’Leary, A. P., & Fletcher, W. (2024). Thinking about Believing: Can Metacognitive Reflection Encourage Belief Updating? *Journal of Intelligence*, 12(5), 47. <https://doi.org/10.3390/jintelligence12050047>
- Osorio T., H., & Reyes M., G. (2023). Decision Making in Moral Judgment Context is Modulated by Individual Metacognition. *Psychological Reports*, 00332941231191067. <https://doi.org/10.1177/00332941231191067>
- Osorio T., H., & Reyes M., G. (2025). Decision Making in Moral Judgment Context is Modulated by Individual Metacognition. *Psychological Reports*, 128(4), 2846–2865. <https://doi.org/10.1177/00332941231191067>
- Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., & Richardson, D. C. (2015). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences*, 112(13), 4170–4175. <https://doi.org/10.1073/pnas.1415250112>
- Pärnamets, P., Richardson, D., & Balkenius, C. (2014). *Modelling moral choice as a diffusion process dependent on visual fixations*. Annual Meeting of the Cognitive Science Society.
- Pauls, I. L., Shuman, E., Van Zomeren, M., Saguy, T., & Halperin, E. (2022). Does crossing a moral line justify collective means? Explaining how a perceived moral violation triggers normative and nonnormative forms of collective action. *European Journal of Social Psychology*, 52(1), 105–123. <https://doi.org/10.1002/ejsp.2818>
- Pescetelli, N., Rees, G., & Bahrami, B. (2016). The perceptual and social components of metacognition. *Journal of Experimental Psychology: General*, 145(8), 949–965. <https://doi.org/10.1037/xge0000180>
- Pilarczyk, J., & Kuniecki, M. (2014). Emotional content of an image attracts attention more than visually salient features in various signal-to-noise ratio conditions. *Journal of Vision*, 14(12), 4–4. <https://doi.org/10.1167/14.12.4>

- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>
- 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. <https://doi.org/10.1002/ejsp.2906>
- Puryear, C., Brady, W. J., Jackson, J. C., Leong, Y. C., & Kteily, N. S. (2025). *Rising Moralization in Social Media Discourse*. PsyArXiv. https://doi.org/10.31234/osf.io/z7p3u_v1
- Qu, C., Bénistant, J., & Dreher, J.-C. (2022). Neurocomputational mechanisms engaged in moral choices and moral learning. *Neuroscience & Biobehavioral Reviews*, 132, 50–60. <https://doi.org/10.1016/j.neubiorev.2021.11.023>
- Qu, C., Hu, Y., Tang, Z., Derrington, E., & Dreher, J.-C. (2020). Neurocomputational mechanisms underlying immoral decisions benefiting self or others. *Social Cognitive and Affective Neuroscience*, 15(2), 135–149. <https://doi.org/10.1093/scan/nsaa029>
- Radke, H. R. M., Kutlaca, M., Siem, B., Wright, S. C., & Becker, J. C. (2020). Beyond Allyship: Motivations for Advantaged Group Members to Engage in Action for Disadvantaged Groups. *Personality and Social Psychology Review*, 24(4), 291–315. <https://doi.org/10.1177/1088868320918698>
- Ramezani, A., Stellar, J. E., Feinberg, M., & Xu, Y. (2026). Historical reconstruction of human moralization with word association and text corpora. *Nature Communications*. <https://doi.org/10.1038/s41467-025-67891-2>
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108. <https://doi.org/10.1037/0033-295X.85.2.59>
- Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, 20(4), 873–922. <https://doi.org/10.1162/neco.2008.12-06-420>
- Redlawsk, D. P. (2002). Hot Cognition or Cool Consideration? Testing the Effects of Motivated Reasoning on Political Decision Making. *The Journal of Politics*, 64(4), 1021–1044. <https://doi.org/10.1111/1468-2508.00161>
- Rekker, R. (2022). Political Polarization Over Factual Beliefs. In J. Strömbäck, Å. Wikforss, K. Glüer, T. Lindholm, & H. Oscarsson, *Knowledge Resistance in High-Choice Information Environments* (1st ed., pp. 222–236). Routledge. <https://doi.org/10.4324/9781003111474-12>
- Reynolds, S. J. (2008). Moral attentiveness: Who pays attention to the moral aspects of life? *Journal of Applied Psychology*, 93(5), 1027–1041. <https://doi.org/10.1037/0021-9010.93.5.1027>

- Rhee, J. J., Schein, C., & Bastian, B. (2019). The what, how, and why of moralization: A review of current definitions, methods, and evidence in moralization research. *Social and Personality Psychology Compass*, 13(12), e12511. <https://doi.org/10.1111/spc3.12511>
- Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018a). Metacognitive Failure as a Feature of Those Holding Radical Beliefs. *Current Biology*, 28(24), 4014-4021.e8. <https://doi.org/10.1016/j.cub.2018.10.053>
- Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018b). Metacognitive Failure as a Feature of Those Holding Radical Beliefs. *Current Biology*, 28(24), 4014-4021.e8. <https://doi.org/10.1016/j.cub.2018.10.053>
- Rottman, J., Young, L., & Kelemen, D. (2017). The impact of testimony on children's moralization of novel actions. *Emotion*, 17(5), 811–827. <https://doi.org/10.1037/emo0000276>
- Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human Metacognition Across Domains: Insights from Individual Differences and Neuroimaging. *Personality Neuroscience*, 1, e17. <https://doi.org/10.1017/pen.2018.16>
- Rozin, P. (1999). The Process of Moralization. *Psychological Science*, 10(3), 218–221. <https://doi.org/10.1111/1467-9280.00139>
- Rozin, P., Markwith, M., & Stoess, C. (1997). Moralization and Becoming a Vegetarian: The Transformation of Preferences Into Values and the Recruitment of Disgust. *Psychological Science*, 8(2), 67–73. <https://doi.org/10.1111/j.1467-9280.1997.tb00685.x>
- Rozin, P., & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review*, 5(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2
- Rozin, P., & Singh, L. (1999). The Moralization of Cigarette Smoking in the United States. *Journal of Consumer Psychology*, 8(3), 321–337. https://doi.org/10.1207/s15327663jcp0803_07
- Rudorf, S., & Hare, T. A. (2014). Interactions between Dorsolateral and Ventromedial Prefrontal Cortex Underlie Context-Dependent Stimulus Valuation in Goal-Directed Choice. *The Journal of Neuroscience*, 34(48), 15988–15996. <https://doi.org/10.1523/JNEUROSCI.3192-14.2014>
- Russell, G. M. (2011). Motives of Heterosexual Allies in Collective Action for Equality. *Journal of Social Issues*, 67(2), 376–393. <https://doi.org/10.1111/j.1540-4560.2011.01703.x>
- Ryan, T. J. (2014). Reconsidering Moral Issues in Politics. *The Journal of Politics*, 76(2), 380–397. <https://doi.org/10.1017/S0022381613001357>
- Ryan, T. J. (2019). Actions versus Consequences in Political Arguments: Insights from Moral Psychology. *The Journal of Politics*, 81(2), 426–440. <https://doi.org/10.1086/701494>

- Said, N., Fischer, H., & Anders, G. (2021). *Contested Science: Individuals with higher metacognitive insight into interpretation of evidence are less likely to polarize* (p. 63719 Bytes) [Dataset]. figshare. <https://doi.org/10.6084/M9.FIGSHARE.13139351>
- Samaha, J., & Postle, B. R. (2017). Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. *Proceedings of the Royal Society B: Biological Sciences*, 284(1867), 20172035. <https://doi.org/10.1098/rspb.2017.2035>
- Saroglou, V., & Craninx, M. (2021). Religious moral righteousness over care: A review and a meta-analysis. *Current Opinion in Psychology*, 40, 79–85. <https://doi.org/10.1016/j.copsyc.2020.09.002>
- Seamans, J. K., & Floresco, S. B. (2022). Event-based control of autonomic and emotional states by the anterior cingulate cortex. *Neuroscience & Biobehavioral Reviews*, 133, 104503. <https://doi.org/10.1016/j.neubiorev.2021.12.026>
- Senko, C., Perry, A. H., & Greiser, M. (2022). Does triggering learners' interest make them overconfident? *Journal of Educational Psychology*, 114(3), 482–497. <https://doi.org/10.1037/edu0000649>
- Sevinc, G., Gurvit, H., & Spreng, R. N. (2017). Salience network engagement with the detection of morally laden information. *Social Cognitive and Affective Neuroscience*, 12(7), 1118–1127. <https://doi.org/10.1093/scan/nsx035>
- Shenhav, A., & Greene, J. D. (2014). Integrative Moral Judgment: Dissociating the Roles of the Amygdala and Ventromedial Prefrontal Cortex. *The Journal of Neuroscience*, 34(13), 4741–4749. <https://doi.org/10.1523/JNEUROSCI.3390-13.2014>
- Skitka, L. J., Bauman, C. W., & Lytle, B. (2009). The Limits of Legitimacy: Morality as a Constraint on Deference to Authority. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1493520>
- Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral Conviction: Another Contributor to Attitude Strength or Something More? *Journal of Personality and Social Psychology*, 88(6), 895–917. <https://doi.org/10.1037/0022-3514.88.6.895>
- Skitka, L. J., Hanson, B. E., Morgan, G. S., & Wisneski, D. C. (2021). The Psychology of Moral Conviction. *Annual Review of Psychology*, 72(1), 347–366. <https://doi.org/10.1146/annurev-psych-063020-030612>
- Skitka, L. J., Hanson, B. E., Washburn, A. N., & Mueller, A. B. (2018). Moral and religious convictions: Are they the same or different things? *PLOS ONE*, 13(6), e0199311. <https://doi.org/10.1371/journal.pone.0199311>
- Skitka, L. J., & Morgan, G. S. (2014). The Social and Political Implications of Moral Conviction. *Political Psychology*, 35(S1), 95–110. <https://doi.org/10.1111/pops.12166>

- Skitka, L. J., & Mullen, E. (2002). The Dark Side of Moral Conviction. *Analyses of Social Issues and Public Policy*, 2(1), 35–41. <https://doi.org/10.1111/j.1530-2415.2002.00024.x>
- Skitka, L. J., Washburn, A. N., & Carsel, T. S. (2015). The psychological foundations and consequences of moral conviction. *Current Opinion in Psychology*, 6, 41–44. <https://doi.org/10.1016/j.copsyc.2015.03.025>
- Skitka, L. J., & Wisneski, D. C. (2011). Moral Conviction and Emotion. *Emotion Review*, 3(3), 328–330. <https://doi.org/10.1177/1754073911402374>
- Soutschek, A., Sauter, M., & Schubert, T. (2015). The Importance of the Lateral Prefrontal Cortex for Strategic Decision Making in the Prisoner’s Dilemma. *Cognitive, Affective, & Behavioral Neuroscience*, 15(4), 854–860. <https://doi.org/10.3758/s13415-015-0372-5>
- Spears, R. (2021). Social Influence and Group Identity. *Annual Review of Psychology*, 72(1), 367–390. <https://doi.org/10.1146/annurev-psych-070620-111818>
- Stanton, J. D., Sebesta, A. J., & Dunlosky, J. (2021). Fostering Metacognition to Support Student Learning and Performance. *CBE—Life Sciences Education*, 20(2), fe3. <https://doi.org/10.1187/cbe.20-12-0289>
- Taylor, C. (2012). *Moralism: A study of a vice*. McGill-Queen’s university press.
- Thaler, M. (2026). Good News Is Not a Sufficient Condition for Motivated Reasoning. *Review of Economics and Statistics*, 1–24. <https://doi.org/10.1162/REST.a.1702>
- Thibodeau, P., Peebles, M. M., Grodner, D. J., & Durgin, F. H. (2015). The Wished-For Always Wins Until the Winner Was Inevitable All Along: Motivated Reasoning and Belief Bias Regulate Emotion During Elections. *Political Psychology*, 36(4), 431–448. <https://doi.org/10.1111/pops.12100>
- Thomas, E. F., Bury, S. M., Louis, W. R., Amiot, C. E., Molenberghs, P., Crane, M. F., & Decety, J. (2019). Vegetarian, vegan, activist, radical: Using latent profile analysis to examine different forms of support for animal welfare. *Group Processes & Intergroup Relations*, 22(6), 836–857. <https://doi.org/10.1177/1368430218824407>
- Tomasello, M. (2016). *A natural history of human morality*. Harvard university press.
- Tomasello, M., & Vaish, A. (2013). Origins of Human Cooperation and Morality. *Annual Review of Psychology*, 64(1), 231–255. <https://doi.org/10.1146/annurev-psych-113011-143812>
- Tusche, A., & Hutcherson, C. A. (2018). Cognitive regulation alters social and dietary choice by changing attribute representations in domain-general and domain-specific brain circuits. *eLife*, 7, e31185. <https://doi.org/10.7554/eLife.31185>

- Tustison, N. J., Avants, B. B., Cook, P. A., Yuanjie Zheng, Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, 29(6), 1310–1320. <https://doi.org/10.1109/TMI.2010.2046908>
- Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, 16(1), 55–61. <https://doi.org/10.1038/nrn3857>
- Ugazio, G., Grueschow, M., Polania, R., Lamm, C., Tobler, P., & Ruff, C. (2022). Neuro-computational foundations of moral preferences. *Social Cognitive and Affective Neuroscience*, 17(3), 253–265. <https://doi.org/10.1093/scan/nsab100>
- 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. <https://doi.org/10.1371/journal.pone.0048693>
- Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or Wrong?: The Brain's Fast Response to Morally Objectionable Statements. *Psychological Science*, 20(9), 1092–1099. <https://doi.org/10.1111/j.1467-9280.2009.02411.x>
- Van Prooijen, J.-W., Krouwel, A. P. M., & Emmer, J. (2018). Ideological Responses to the EU Refugee Crisis: The Left, the Right, and the Extremes. *Social Psychological and Personality Science*, 9(2), 143–150. <https://doi.org/10.1177/1948550617731501>
- 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. <https://doi.org/10.1111/j.2044-8309.2010.02000.x>
- Van Zomeren, M., Postmes, T., Spears, R., & Bettache, K. (2011). Can moral convictions motivate the advantaged to challenge social inequality?: Extending the social identity model of collective action. *Group Processes & Intergroup Relations*, 14(5), 735–753. <https://doi.org/10.1177/1368430210395637>
- Wang, S., Yu, R., Tyszka, J. M., Zhen, S., Kovach, C., Sun, S., Huang, Y., Hurlemann, R., Ross, I. B., Chung, J. M., Mamelak, A. N., Adolphs, R., & Rutishauser, U. (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. *Nature Communications*, 8(1), 14821. <https://doi.org/10.1038/ncomms14821>
- Westen, D., Blagov, P. S., Harenski, K., Kilts, C., & Hamann, S. (2006). Neural Bases of Motivated Reasoning: An fMRI Study of Emotional Constraints on Partisan Political Judgment in the 2004 U.S. Presidential Election. *Journal of Cognitive Neuroscience*, 18(11), 1947–1958. <https://doi.org/10.1162/jocn.2006.18.11.1947>
- Westra, E. (2021). Virtue Signaling and Moral Progress. *Philosophy & Public Affairs*, 49(2), 156–178. <https://doi.org/10.1111/papa.12187>

- White, S. F., Zhao, H., Leong, K. K., Smetana, J. G., Nucci, L. P., & Blair, R. J. R. (2017). Neural correlates of conventional and harm/welfare-based moral decision-making. *Cognitive, Affective, & Behavioral Neuroscience*, 17(6), 1114–1128. <https://doi.org/10.3758/s13415-017-0536-6>
- Wills, J., FeldmanHall, O., NYU PROSPEC Collaboration, Meager, M. R., Van Bavel, J. J., NYU PROSPEC Collaboration, Blackmon, K., Devinsky, O., Doyle, W. K., Luciano, D. J., Kuzniecky, R. I., Nadkarni, S. S., Vazquez, B., Najjar, S., Geller, E., Golfinos, J. G., Placantonakis, D. G., Friedman, D., Wisoff, J. H., & Samadani, U. (2018). Dissociable contributions of the prefrontal cortex in group-based cooperation. *Social Cognitive and Affective Neuroscience*, 13(4), 349–356. <https://doi.org/10.1093/scan/nsy023>
- Wisneski, D. C., & Skitka, L. J. (2017). Moralization Through Moral Shock: Exploring Emotional Antecedents to Moral Conviction. *Personality and Social Psychology Bulletin*, 43(2), 139–150. <https://doi.org/10.1177/0146167216676479>
- Workman, C. I., Yoder, K. J., & Decety, J. (2020). The Dark Side of Morality – Neural Mechanisms Underpinning Moral Convictions and Support for Violence. *AJOB Neuroscience*, 11(4), 269–284. <https://doi.org/10.1080/21507740.2020.1811798>
- Wright, J. C. (2012). Children’s and adolescents’ tolerance for divergent beliefs: Exploring the cognitive and affective dimensions of moral conviction in our youth. *British Journal of Developmental Psychology*, 30(4), 493–510. <https://doi.org/10.1111/j.2044-835X.2011.02058.x>
- Wright, J. C., Cullum, J., & Schwab, N. (2008). The Cognitive and Affective Dimensions of Moral Conviction: Implications for Attitudinal and Behavioral Measures of Interpersonal Tolerance. *Personality and Social Psychology Bulletin*, 34(11), 1461–1476. <https://doi.org/10.1177/0146167208322557>
- Wright, J. C., & Pölzler, T. (2022). Should morality be abolished? An empirical challenge to the argument from intolerance. *Philosophical Psychology*, 35(3), 350–385. <https://doi.org/10.1080/09515089.2021.1983160>
- Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS ONE*, 8(7), e68910. <https://doi.org/10.1371/journal.pone.0068910>
- Ye, Q., Zou, F., Lau, H., Hu, Y., & Kwok, S. C. (2018). Causal Evidence for Mnemonic Metacognition in Human Precuneus. *The Journal of Neuroscience*, 38(28), 6379–6387. <https://doi.org/10.1523/JNEUROSCI.0660-18.2018>
- Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1310–1321. <https://doi.org/10.1098/rstb.2011.0416>
- Yin, D., Mitra, S., & Zhang, H. (2016). Research Note—When Do Consumers Value Positive vs. Negative Reviews? An Empirical Investigation of Confirmation Bias in Online Word of

- Mouth. *Information Systems Research*, 27(1), 131–144.
<https://doi.org/10.1287/isre.2015.0617>
- Yoder, K. J., & Decety, J. (2018). The neuroscience of morality and social decision-making. *Psychology, Crime & Law*, 24(3), 279–295.
<https://doi.org/10.1080/1068316X.2017.1414817>
- Yoder, K. J., & Decety, J. (2022). Moral conviction and metacognitive ability shape multiple stages of information processing during social decision-making. *Cortex*, 151, 162–175.
<https://doi.org/10.1016/j.cortex.2022.03.008>
- Zaal, M. P., Saab, R., O'Brien, K., Jeffries, C., Barreto, M., & Van Laar, C. (2017). You're either with us or against us! Moral conviction determines how the politicized distinguish friend from foe. *Group Processes & Intergroup Relations*, 20(4), 519–539.
<https://doi.org/10.1177/1368430215615682>
- Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), 45–57. <https://doi.org/10.1109/42.906424>
- Zmigrod, L., Rentfrow, P. J., & Robbins, T. W. (2020). The partisan mind: Is extreme political partisanship related to cognitive inflexibility? *Journal of Experimental Psychology: General*, 149(3), 407–418. <https://doi.org/10.1037/xge0000661>