UNIVERSITY OF CALIFORNIA

Santa Barbara

Theories of Morality and Media: Examining Representations of Moral Cognition and the Modulating Effects of Moral Domain Sensitivity

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Communication

by

Sungbin Youk

Committee in charge:

Professor René Weber, Chair

Professor Daniel Linz

Professor Jiaying Liu

March 2024

The dissertation of Sungbin Youk is approved.

Daniel Linz

Jiaying Liu

René Weber, Committee Chair

February 2024

Theories of Morality and Media: Examining Representations of Moral Cognition and the Modulating Effects of Moral Domain Sensitivity

ACKNOWLEDGEMENTS

I am immensely grateful for the unwavering support of my family, friends, colleagues, and the entire department, without whom this work would not have been possible. I extend my heartfelt appreciation to everyone who played a role in my growth as a social scientist throughout my Ph.D. journey.

Before embarking on my Ph.D. journey, I fondly recall discussing the essential characteristics of a successful Ph.D. student with my mentors, Timothy (Tim) R. Levine and Hee Sun Park, during a memorable lunch get-together. To my surprise, Tim emphasized the significance of mental health. I now fully comprehend the gravity of this perspective. I want to express special thanks to my lifelong partner, Hyun Ji Lee, who not only supported me but also steadfastly believed in my academic journey, consistently prioritizing my mental well-being. My gratitude extends to my in-laws and parents for their unwavering encouragement.

As the conclusion of my Ph.D. marks the beginning of a new chapter in scholarship, I extend my thanks to my committee members, my advisor, and the members of the Media Neuroscience Lab. I look forward to continued collaboration for scientific improvement, with a commitment to mutually inspiring each other academically.

VITA OF SUNGBIN YOUK
February 2024

EDUCATION

Ph.D., Communication *Jan 2021 - Feb 2024*
University of California, Santa Barbara
Advisor: René Weber

M.A., Media & Communication *Sep 2018 - Aug 2020*
Korea University
Advisor: Hee Sun Park

B.A., International Studies *Mar 2012 - Aug 2018*
Media & Communication (Double Major)
Korea University

REFEREED PUBLICATIONS

Corresponding Author

Jeong, D., & **Youk, S.*** (2023). Refining esports: A quantitative cartography of esports literature. *Entertainment Computing, 47*, article 100597. https://doi.org/10.1016/j.entcom.2023.100597

Youk, S., Malik, M., Chen, Y., Hopp, F. R., & Weber, R. (2023). Measures of intrinsic argument strength: A computational, large-scale analysis of effective persuasion in real-world debates. *Communication Methods and Measures*. https://doi.org/10.1080/19312458.2023.2230866

Chen, Y., Youk, S., Wang, P. T., Pinti, P., Weber, R. (2023). A calculus of probability or belief? Neural underpinnings of social decision-making in a card game. *Neuropsychologia*, 188, article 108635. https://doi.org/10.1016/j.neuropsychologia.2023.108635

Youk, S., & Park, H. S. (2023). Who is (communicatively more) responsible behind the wheel? Applying the theory of communicative responsibility to TAM in the context of using navigation technology. *Human-Machine Communication, 6*, 205-230. https://doi.org/10.30658/hmc.6.11

Youk, S., & Park, H. S. (2023). Why scapegoating can ruin an apology: The mediated-moderation model of appropriate crisis response messages in the context of South Korea. *Frontiers in Psychology*, *13,* article 1082152. https://doi.org/10.3389/fpsyg.2022.1082152

Im, W. J., Youk, S., & Park, H. S. (2021). Apologies combined with other crisis response strategies: Do the fulfillment of individuals' needs to be heard and the timing of response message affect apology appropriateness? *Public Relations Review, 47*(1), article 102002. https://doi.org/10.1016/j.pubrev.2020.102002

Lim, J. I., Youk, S., & Park, H. S. (2019). How to communicate in the peer review process: A gentle guideline for novice researchers. *Asian Communication Research, 16*(3), 179-202. https://doi.org/10.20879/acr.2019.16.3.179

Youk, S., & Park, H. S. (2019). Where and what do they publish? Editors' and editorial board members' affiliated institutions and the citation counts of their endogenous publications in the field of communication. *Scientometrics, 120*(3), 1237-1260. https://doi.org/10.1007/s11192-019-03169-x

SELECTED CONFERENCE PRESENTATIONS

Lee, H., Youk, S., Lee, Y. E., Malik, M., Weber, R. (2023). *Viewer engagement with Autonomous Sensory Meridian Response Videos (ASMR): Is there neurological evidence for their therapeutic relevance in stressed and lonely individuals?* [Poster presentation]. The 14th Annual Meeting of Social & Affective Neuroscience Society, Santa Barbara, CA, USA.

Malik, M., Youk, S., & Weber, R. (2023). *Evaluating the structural position, cinematographic representation, & emotional portrayal of female characters in international feature films.* [Paper presentation]. The 73rd Annual Convention of the International Communication Association, Toronto, Canada. ***TOP PAPER**

Woodman, K., Youk, S., Wheeler, B., & Weber, R. (2023). *Video gaming, a protective factor or a gateway drug for early adolescent substance use.* [Paper presentation]. The 73rd Annual Convention of the International Communication Association, Toronto, Canada.

Youk, S., Han, J., & Park, H. S. (2022). *How far should the franchise apologize? Integration of product and value related crisis types to situational crisis communication theory* [Paper presentation]. The 108th Annual Convention of the National Communication Association. New Orleans, LA, United States.

Youk, S., Wang, H., Malik, M., & Weber, R. (2022). *Moral themes in lyrics of popular rap songs in comparison to other music genres: An application of the Model of Intuitive Morality and Exemplars.* [Poster presentation]. The Annual Meeting of Media & Morality. Davis, CA, United States.

Jeong, D., & Youk, S. (2021). *Who is Researching what in eSports Literature? A cartography of authors and keywords for eSports studies using network analysis* [Paper presentation]. The 107th Annual Convention of the National Communication Association. Seattle, WA, United States. ***TOP STUDENT PAPER**

Youk, S., Park, H. S. (2021). *Who is responsible behind the wheel? Integrating the theory of communicative responsibility and TAM in the light of the CASA paradigm* [Paper presentation]. The 71st Annual Convention of the International Communication Association. Virtual.

Youk, S., Jeong, D., & Park, H. S. (2020). *Examining the moderated mediation effects of apologetic crisis response strategies on the organization's overall reputation by applying theory of communicative responsibility* [Paper presentation]. The 70th Annual Convention of the International Communication Association, Gold Coast, Australia. ***TOP PAPER**

Youk, S., Park, H. S., Ryu, J. Y., Lim, J. I., & Han, J.-H. (2019). *Geographical Location of Institutional Affiliation and Publication Types of Editors and Editorial Board Members in the Field of Communication* [Paper presentation]. The 69th Annual Convention of the International Communication Association, Washington, DC, USA.

Im, W. J., Park, H. S., Youk, S., Ryu, J. Y., & Oh, Y. J. (2018). *Will you accept our apology? Evaluation of timing of announcement, type of apology, and fulfillment of individual's needs to be heard as a crisis management strategy* [Paper presentation]. The 104th Annual Convention of the National Communication Association, Salt Lake City, UT, USA.

TEACHING EXPERIENCE

Teaching Associate

Summer Research Academies, University of California, Santa Barbara
INT 93LS Science of Persuasion — *Summer 2022*
Department of Communication, University of California, Santa Barbara
COMM1 Introduction to Communication — *Summer 2022*
COMM1 Introduction to Communication — *Summer 2021*

Teaching Assistance

Summer Research Academies, University of California, Santa Barbara
INT 93LS Moral Mining — *Summer 2023*
Department of Communication, University of California, Santa Barbara
COMM 104 Evolution & Human Communication — *Fall Quarter 2023*
COMM 1 Introduction to Communication — *Summer 2023*
COMM 160MH Mental Health Communication — *Spring Quarter 2023*
COMM 87 Statistical Analysis for Communication — *Winter Quarter 2023*
COMM 89 Communication Theory — *Spring Quarter 2022*
COMM 89 Communication Theory — *Winter Quarter 2022*
COMM 1 Introduction to Communication — *Fall Quarter 2021*
COMM 1 Introduction to Communication — *Spring Quarter 2021*
COMM 1 Introduction to Communication — *Winter Quarter 2021*
School of Media & Communication, Korea University
JMCO264 Interpersonal Communication — *Fall Semester 2019*
JMCO264 Interpersonal Communication — *Spring Semester 2019*
JMCO264 Interpersonal Communication — *Fall Semester 2018*
JMC579 Data Analysis I — *Spring Semester 2018*

MENTORSHIP

Rohan Sontakke, Dublin High School (Student Paper presented at 2023 NCA)
Mihir S. Arya, University of Michigan (Student Paper presented at 2023 NCA)
Kevin Hao Chuan Wang, University of Chicago (Poster presented at 2022 Media & Morality)
Junhyung Han, University of Illinois Urbana-Champaign (Paper presented at 2022 NCA)
Dong Wook Jung, Pennsylvania State University (Paper presented at 2021 NCA)

AWARDS, CERTIFICATES, HONORS, SCHOLARSHIPS, & FELLOWSHIP

Outstanding Graduate Student Service Award 2023
Department of Communication, University of California, Santa Barbara

Top Paper Award *2023*
Computational Methods Division, International Communication Association

Course Design for Equity Certificate *2022*
CITRAL Community of Practice, University of California, Santa Barbara

Edwin Schoell Award for Excellence in Teaching *2022*
Department of Communication, University of California, Santa Barbara

Teaching Certificate *2021*
Summer Teaching Institute for Associates, University of California, Santa Barbara

Top Student Paper Award *2021*
Game Studies Division, International Communication Association

Top Paper Award *2020*
Korean American Communication Association Division, International Communication Association

Teaching Fellow Scholarship *2019*
School of Media & Communication, Korea University

General Scholarship *2019*
School of Media & Communication, Korea University

Research Award *2018*
Ministry of Unification, South Korea

Veritas Program Scholarship *2018*
School of Media & Communication, Korea University

Honors Scholarship *2012, 2013, 2014*
College of International Studies, Korea University

SERVICE

Ph.D Representative of Graduate Student Advisory Committee to the Chair *2023*
Department of Communication, University of California, Santa Barbara

Reviewer *2023*
Computational Communication Research, Human-Machine Communication

Lab Manager *2022, 2023*
Media Neuroscience Lab, Department of Communication, University of California, Santa Barbara

Graduate Student Representative of Job Search Committee *2022*
Department of Communication, University of California, Santa Barbara

Data Science Consultation *2022*
Center For Information Technology and Society, University of California, Santa Barbara

IT Coordinator *2021*
Media Neuroscience Lab, Department of Communication, University of California, Santa Barbara

Graduate Student Volunteer for Open House *2021, 2023*
Department of Communication, University of California, Santa Barbara

ABSTRACT

Theories of Morality and Media: Examining Representations of Moral Cognition and the Modulating Effects of Moral Domain Sensitivity

by

Sungbin Youk

Moral judgment, a fundamental aspect of human behavior, involves evaluative distinctions between actions deemed morally "good" or "bad." This dissertation delves into the cognitive and neurological processes underpinning moral judgment by examining two influential theoretical frameworks: Moral Foundations Theory (MFT) and Morality-as-Cooperation (MAC). MFT posits innate moral foundations (e.g., harm, fairness, loyalty) as the basis of morality, while MAC focuses on moral elements (e.g., helping kin, fairness) emerging from cooperation challenges. Despite the pervasive influence of morality research on communication and media content, particularly in areas like moralizing language on social media and persuasion strategies using moral framing, a comprehensive comparison of the neurological underpinnings across these theories has been lacking in the literature. The dissertation reveals neural networks related to the theory of mind to be associated with moral cognition across the two theories. However, it also reveals that the neural representation for each moral foundation and element exhibits distinct patterns. The factorization of neural representations of MFT and MAC provides robust evidence for the theoretical foundations of

both frameworks, emphasizing the theoretical overlap of moral domains across the two theories. Moreover, different survey measures for moral sensitivity yield variations in predicting the neural representation of moral cognition, illustrating how a neuroscientific approach can offer additional validations of survey measures.This dissertation offers a nuanced understanding of moral cognition, revealing the complexity of neural representations associated with different moral foundations and elements from two competing theories of moral cognition.

TABLE OF CONTENTS

I. Introduction

Moral judgments represent a fundamental aspect of human behavior, encompassing the evaluative distinctions between actions deemed morally "good" or "bad" (Cheng et al., 2021). This intrinsic aspect not only significantly influences day-to-day behaviors but also affects multiple facets of communication. It is evident in various media content, encompassing (but not limited to) social media and feature films. The judgments in response to such portrayals shape individual values, subsequently wielding influence over societal morality (Tamborini, 2011). While a moral verdict may often feel instinctive, the underlying cognitive processes are far from simple.

Research on morality is indispensable as it permeates various aspects of communication and media content. A notable example is the pervasive use of moralizing language in polarizing social media content (Brady et al., 2017), where moral appeals frequently heighten ideological divides and contribute to the formation of echo chambers. Furthermore, the moral framing employed in news narratives significantly influences audience perception (Fulgoni et al., 2016), as the presentation of information through distinct moral lenses can sway public opinion and exacerbate societal tensions. Additionally, media entertainment narratives, ranging from television shows to films, can be more comprehensively understood by examining their content (Lewis et al., 2014). This exploration sheds light on the moral values embedded within the narratives and their potential impact on viewers' enjoyment and appraisal.

While numerous studies in communication have applied a moral framework in their research, less is understood about its foundation: the underlying mechanisms of moral

cognition. To advance our understanding of moral cognition and facilitate the integration of moral perspectives in communication research, we take a neuroscientific perspective into further examining two prominent theoretical frameworks: Moral Foundations Theory (MFT) and Morality-as-Cooperation (MAC). MFT posits that morality is underpinned by a set of innate moral foundations, each serving as an evolutionary adaptation (Graham et al., 2009). These foundations include care, fairness, loyalty, authority, and purity. On the other hand, MAC argues that morality arises from a collection of biological and cultural solutions to the recurrent problems of cooperation in human social life (Curry et al., 2019). This theory introduces seven types of cooperation (e.g., helping kin and helping one's group), corresponding to seven types of morality. While prior literature has made significant strides in identifying neural substrates associated with moral judgment and dissecting the moral foundations of MFT (Hopp et al., 2023), a notable gap exists in comparing neurological underpinnings between MFT and MAC. Both theories converge in their acknowledgment of biological and evolutionary origins but diverge when it comes to moral domains. Examining the two theories from a neurological standpoint offers the prospect of a comprehensive understanding of the cognitive processes involved in moral judgment.

The fundamental objective of this dissertation proposal is to explore how the well-established theory-driven moral domains, referred to as foundations in MFT and elements in MAC, translate into neural domains. We aim to decipher how different regions or sets of regions in the brain engage during moral judgment tasks when invoking these theory-driven moral foundations and elements. This endeavor not only enhances our understanding of moral cognition but also contributes to theoretical advancements in moral research. Furthermore, we delve into how individual disparities in moral intuition salience

manifest in distinct patterns of brain activity. To accomplish this, this dissertation proposal is organized as follows. Firstly, we present an overview highlighting the significance of morality, specifically delving into moral psychology within the realm of communication. Subsequently, we discuss the theoretical underpinnings of the two focal theories, elucidating how each theory conceptualizes morality, their respective distinctions, and points of intersection. In the third section, the dissertation proposal introduces neurological evidence concerning morality, utilizing two contrasting neuroscientific approaches. Moreover, we engage in a discourse surrounding the modularity of morality by synthesizing theoretical arguments and neuroscientific evidence. Additionally, we explore the intricate interplay between individual differences in moral domain silence and neural activity during moral judgment.

II. Literature Review

A. Relevance of Morality in Media and Communication

Considering how moral values are the underlying fundamentals of people's attitudes and identities (Aquino & Reed, 2002; Strohminger & Nichols, 2014), it is not surprising that scholars have extensively explored the relevance of morality in diverse communication contexts, including but not limited to persuasion, interpersonal communication, social media, news, and entertainment media. The exploration of morality within these communication realms is crucial, as it unveils the nuanced ways in which moral considerations intricately shape the construction and interpretation of messages. By delving into the interplay of morality within diverse communication contexts, we gain valuable insights into how moral values contribute to the fabric of communicative processes, influencing not only the

transmission but also the reception and understanding of messages in our interconnected and media-saturated society.

In the realm of persuasion literature, the research underscores that aligning persuasive messages with the moral beliefs of the audience enhances their effectiveness, fostering a connection that transcends mere logical appeal (Aramovich et al., 2012; Feinberg & Willer, 2015; Koleva et al., 2012; Luttrell et al., 2017; Ryan, 2017; Weber et al., 2015). In other words, the persuasive power of arguments is notably heightened when they resonate with individuals' deeply held moral values. Even arguments initially met with resistance can be rendered more compelling through a strategic emphasis on the moral values salient to the audience, a phenomenon known as moral reframing (Feinberg & Willer, 2019). This nuanced approach acknowledges the capacity to reshape perspectives by reframing arguments within the moral context that holds significance for the audience. Additionally, recent research by Youk et al. (2023) delves into the dynamics of online debate platforms, investigating how moral values and the similarity of moral content can amplify the persuasive strength of arguments. This exploration not only sheds light on the evolving landscape of persuasion in the digital age but also underscores the enduring role of morality in shaping the effectiveness of persuasive communication strategies.

Interpersonal communication within the family unit is paramount for the socialization and negotiation of morality. Ochs and Kremer-Sadlik (2007) emphasize that a universal function of the family is to cultivate the thinking and emotions of children in alignment with moral ideals. As early as the age of three, children begin to internalize moral values and rules through their interactions with caregivers (Buchsbaum & Emde, 1990), solidifying the family's status as a pivotal influence on moral upbringing (Lollis et al., 1996; White &

Matawie, 2004). Beyond the developmental aspect, the family functions as a crucible for negotiating and communicating the moral compass. Individuals can openly and securely challenge one another's moral beliefs and judgments within familial boundaries. For instance, family discussions may provide a platform for reflecting on the moral values of a child's actions, fostering an environment where responsibility and accountability are explored during dinner conversations (Sterponi, 2003). Instances of conflict over moral transgressions are common, as parents and family members feel a sense of responsibility to guide children toward a morally upright path. Parental intervention in sibling conflicts, for example, serves as a proactive measure to impart moral principles (Lollis et al., 1996).

In the realm of social media, the discourse transcends mere information sharing and extends into the realm of moral inducements. Online social networks have evolved into pervasive platforms for discussing moral and political ideas, with implications for disseminating messages. Research by Brady et al. (2017) reveals a noteworthy association between the use of moral-emotional language in political messages and a substantial increase in their diffusion. The strategic use of moralizing language is also used by political elites (Brady et al., 2019). A recent review of scholarly work identified approximately 80 publications providing an overview of moral research in social media across business, psychology, and communication journals (Neumann & Rhodes, 2023). However, the prevalence of atheoretical perspectives in these research publications is an intriguing observation. Furthermore, scholarly investigations are extending beyond the confines of U.S. populations, as evidenced by Singh et al. (2021), underscoring the global nature of moral language usage in social media.

In the domain of news media communication, the role of morality is multifaceted and

influential, shaping both the emotional responses of audiences and the framing of news issues. Research by Bruns (2022) demonstrates the potent impact of journalistic reports on moral violations, emphasizing their ability to evoke strong emotional reactions and significantly affect memory retention. As highlighted by Fulgoni et al. (2016), framing news issues is a dynamic process influenced by different notions of morality employed by news sources. This variability in framing not only shapes current perceptions but also forecasts future news frames and events, as demonstrated in the study by Hopp et al. (2020). Moreover, detecting fake news is intricately tied to moral considerations, with Carvalho et al. (2020) revealing the utility of moralizing language as a discerning factor in identifying deceptive information. This body of research collectively underscores the pervasive influence of morality in news media communication, impacting not only the emotional landscape of audiences but also the broader framing and dynamics of news coverage.

Both theoretically and practically, the realm of media entertainment has seen significant progress through integrating moral perspectives. One noteworthy theoretical advancement is Tamborini's Model of Intuitive Morality Exemplars (MIME), articulated in 2011 and 2013. This framework delves into the intricacies of group-based values, media selection, and the subsequent effects of media consumption, operating at both individual and societal levels. Unlike previous paradigms that primarily viewed media as a reflection of existing cultural norms, MIME goes beyond and discusses how media exposure molds and influences the moral compass of individuals. It also explores how media serves as a vehicle for reinforcing shared moral values through a reciprocal relationship between media and morality.

Research within the MIME framework has uncovered distinct patterns in narratives

featuring conflicts in moral intuitions. For instance, narratives with moral conflicts elicit deeper appreciation and slower appraisal than narratives without moral intuition conflicts (Lewis et al., 2014). A quasi-experimental study conducted over eight weeks, exposing selected participants to an online soap opera, consistently aligns with MIME's postulation that entertainment media can shape moral judgments (Eden et al., 2014). By examining these findings, we gain insight into how entertainment media, guided by moral frameworks like MIME, not only reflects but actively shapes and influences moral judgments, contributing to the complex interplay between media narratives and individuals' moral perspectives.

To further comprehend the intricacies of moral cognition, it is essential to delve into the underlying neurophysiological mechanisms. This includes exploring the intricate interplay between exposure to moralizing stimuli and cognitive processes (Eden et al., 2014). Despite the growing recognition of the importance of neuroscientific approaches within communication scholarship (Hopp et al., 2023; Weber et al., 2018), there remains a notable gap that this dissertation seeks to address. By venturing into neuroscientific approaches to investigate moral cognition, this research aims to scrutinize the processes underpinning moral judgments, unravel the fundamental mechanisms at play, and provide a potential avenue for decoding cognitive processes related to morality. The ultimate goal is to facilitate causal predictions of moral stimuli and identify the neural correlates specifically linked to these stimuli, thereby advancing the field of communication research and offering a more comprehensive understanding of the intricate relationship between media content and moral cognition.

B. Two Theories of Morality

1. Moral Foundations Theory

One prominent theory in moral psychology that elucidates the development and distinct moral domains is the Moral Foundations Theory (MFT). Rooted in the integration of psychological, developmental, and evolutionary perspectives, MFT posits that every individual possesses an inherent sense of moral knowledge derived from recurrent social challenges and opportunities encountered by the species over extended periods of time (Graham et al., 2013). These foundations represent the fundamental components of morality by compiling past experiences and emotions into intuitive mental constructs (Tamborini, 2011). MFT also suggests that this innate moral knowledge is not static and undergoes modification through cultural learning, allowing specific moral foundations to gain prominence due to various external and internal influences (Graham et al., 2009). In other words, MFT elucidates both the variances and universal aspects of moral judgments across diverse cultures. Thus, MFT centers around four key claims: (a) evolutionary processes have crafted an initial framework for the moral mind; (b) the initial configuration of the moral mind is subject to development in response to cultural influences; (c) when witnessing moral transgressions or moral behaviors, intuitions precede justifications; and (d) morality is not monolithic but pluralistic, comprising multiple foundations (often referred to as modules).

While individuals may differ in the extent to which they endorse various moral foundations, an abundance of research on MFT has extensively examined and validated five moral domains that are consistent across cultures (Atari et al., 2023; Graham et al., 2013; Haidt & Joseph, 2004): (1) Care/Harm relates to humans' evolutionary capacity to feel and

empathize with the pain of others, underlying concepts of kindness, gentleness, and nurturance; (2) Fairness/Cheating is associated with evolutionary mechanisms that make humans sensitive to reciprocity and equality, evoking intuitions of justice, rights, and autonomy; (3) Loyalty/Betrayal is influenced by evolutionary mechanisms that emphasize reciprocity and evoke intuitions related to justice, rights, and autonomy; (4) Authority/Subversion is shaped by society's long-standing acceptance and establishment of hierarchical social interactions, emphasizing the importance of obeying legitimate authority and respecting traditions; and (5) Sanctity/Degradation is linked to the psychology of disgust and contamination in human beings, underlying the value of safeguarding oneself against contamination. While bodily purity is especially pronounced in contexts of cleanliness, this foundation also extends into the spiritual realm.

The proponents of MFT emphasize the concept of "moral pluralism," asserting that morality comprises a plethora of foundations rather than a fixed number of foundations (Graham et al., 2013). As Haidt and Joseph (2011) stated, the proposed foundations serve as “a starting point, not an exhaustive list” (p. 2117). The five foundations, which have undergone extensive empirical investigation, represent a modification from the original four foundations (Haidt & Joseph, 2004). Ongoing discussions revolve around identifying new moral foundations and adjusting existing ones. For instance, proportionality (Medin et al., 2010; Skurka et al., 2020), liberty (Iyer et al., 2012), honor (Atari, Graham, et al., 2020), ownership (Atari & Haidt, 2023), modesty (Suhler & Churchland, 2011), and the division of loyalty into group-level and country-level (Zakharin & Bates, 2023) are subjects of ongoing research and consideration within the framework of MFT.

This dissertation proposal is centered on six distinct moral foundations: care, loyalty,

authority, sanctity, equality, and proportionality. The division of fairness into equality and proportionality challenges the notion that fairness can be reduced to a singular conceptualization of resource distribution (Rai, 2018). Equality entails promoting an equitable balance akin to reciprocity in social relationships, emphasizing equal treatment, equal participation, equal opportunities, and equitable resource allocation. Proportionality, often called equity, focuses on ensuring that rewards and punishments are proportionate to individual costs, contributions, effort, merit, or culpability in social interactions (Rai & Fiske, 2011). Empirical research by Atari et al. (2023) conducted across 25 populations demonstrates the dual dimensionality of fairness. Their findings suggest that an individual's low scores on proportionality do not necessarily imply a lack of concern for equality; these two constructs do not represent opposing ends of a single spectrum. These findings have been replicated, reinforcing the existence of the six-foundation structure (Zakharin & Bates, 2023).

2. Morality-as-Cooperation

Another theoretical framework that adopts a moral pluralistic perspective to identify universal moral rules is the Morality-as-Cooperation (MAC) theory proposed by Curry (2016). MAC posits that morality comprises a collection of biological and cultural solutions to the recurrent challenges of cooperation in human social life (Curry, Mullins, et al., 2019). This theory employs non-zero-sum game theory to delineate seven distinct types of cooperation: aiding kin, supporting one's group, reciprocating, displaying courage, deferring to authority, resolving resource disputes, and honoring prior possession. These cooperative behaviors correspond to seven types of morality: (1) family values, (2) group loyalty, (3)

reciprocity, (4) bravery (also called heroism), (5) respect (also called deference), (6) fairness, and (7) property rights.

These seven facets of morality encompass the following: (1) Family values are linked to resource allocation for one's kin, involving cooperative behaviors like caring for offspring, assisting family members, and avoiding inbreeding; (2) Group loyalty pertains to coordinated activities for mutual benefit and includes cooperative behaviors such as forming friendships, engaging in collaborative ventures, supporting coalitions and alliances, and adhering to local customs; (3) Reciprocity is rooted in social exchange and conditional cooperation, encompassing behaviors like trust, repaying favors, seeking retribution, expressing gratitude, and making amends; (4) Bravery embodies heroic virtues, including fortitude, skill, and wit; (5) Respect involves displays of submission, humility, deference, and obedience; (6) Fairness relates to the equitable resolution of disputes over resources, encompassing notions of both equality and proportionality; (7) Property rights encompass respect for prior possessions, which is regarded as morally commendable. Empirical evidence provided by Curry et al. (2019) demonstrates that these seven types of cooperative behaviors are universally considered morally virtuous across 60 societies, lending support to the existence of these seven moral domains.

3. Similarities and Differences Across Two Theories

MAC and MFT share some commonalities. First, they both incorporate evolutionary and psychological perspectives into their frameworks. Second, they embrace a moral pluralistic viewpoint, suggesting that domains of morality can extend beyond their existing foundations and types. Third, while their proposed foundations and types of cooperation are

universally acknowledged, their salience varies among individuals and societies, contingent upon diverse external influences, whether cultural or problem-oriented.

The disparities between these theories pertain to how they establish the foundations and types of cooperation, which can be collectively referred to as moral domains. MAC finds its roots in a unified theoretical foundation, especially non-zero-sum game theory, whereas MFT adopts an ad hoc approach (Curry, Mullins, et al., 2019; Haidt & Joseph, 2011). Consequently, Curry (2016) argues that MFT lacks an underlying theory, thereby rendering it unable to make systematic predictions about the nature of morality. As for MAC, a moral domain should adhere to a cooperative principle (Curry et al., 2019). As proposed by MFT, the moral domains of purity and care lack specific types of cooperation. Moreover, kin altruism, reciprocal altruism, hawkish dominance displays, and property rights, which should be recognized as moral domains, remain unaddressed within MFT. Hence, only three moral domains overlap: loyalty (group loyalty in MAC), equality/proportionality (fairness in MAC), and authority (respect in MAC).

The theoretical distinctions between these two moral pluralist approaches persist in understanding and classifying morality. For example, Haidt and Joseph (2004) might argue that purity is integral to social dynamics, criticizing Curry et al.'s (2019) understanding that avoiding impurity, disease, and unclean behavior is not cooperative. The fear of unfamiliar out-groups may, in fact, stem from concerns over the pathogens they might carry and further disseminate to the in-group (Murray & Schaller, 2016), suggesting that violations of purity can hinder cooperative behaviors. Furthermore, Atari and Haidt (2023) have recently posited that ownership meets the criteria to be considered a moral foundation, incorporating the domain of property rights into the MFT framework.

This dissertation proposal aims to elucidate the extent to which the similarities and distinctions between these two theories can be comprehended through neurological evidence. Given that both theories are grounded in evolutionary psychology, an exploration of moral judgment across various moral domains in response to moral transgressions should be reflected in biological and cognitive evidence. Additionally, considering that both theories discuss individual differences in the salience of these moral domains, this study seeks to examine to what extent these variations account for moral judgment processes as manifested in the brain.

C. Neurological Evidence of Morality

Neuroscientific investigations, encompassing brain lesion case studies, neuroimaging techniques, neurochemical modulation, and insights from evolutionary psychology, have provided empirical support for the neurobiological underpinnings of morality (Moll et al., 2003; Rueda, 2021). Among these methodologies, neuroimaging technology, specifically functional magnetic resonance imaging (fMRI), has significantly advanced our understanding of moral cognition by offering insights into the neural representations and mechanisms at play. The existing body of literature on the neuroimaging of morality emphasizes the compartmentalization of moral cognition through a domain-general approach instead of a domain-specific perspective.

1. Domain-Specific Approach

In the earlier phases of research into the neuroscience of morality, a domain-specific approach prevailed, with a primary focus on pinpointing specific brain regions that uniquely contribute to moral cognition. Brain lesion case studies, for instance, compared individuals'

moral judgments before and after suffering brain impairments. These investigations revealed that lesions in regions such as the frontal cortex and select subcortical nuclei, including the amygdala and ventromedial hypothalamus, induced changes in moral appraisal while leaving other cognitive functions intact (Baez et al., 2014; Ciaramelli et al., 2007; Martins et al., 2012; Moll et al., 2003).

Neuroimaging studies employing a domain-specific approach have aimed to unveil selective neural responses to moral stimuli by meticulously designing controlled comparisons between what are presumed to be moral and non-moral stimuli. For instance, early neuroscientific investigations into moral cognition sought to contrast brain activation patterns when participants engaged in a straightforward moral judgment task (e.g., rating the immorality of statements like "they hung an innocent") versus a non-moral judgment task (e.g., assessing the veracity of statements like "stones are made of water" see Moll et al., 2003). These studies consistently revealed that regions such as the medial frontal gyrus, anterior temporal cortex, left angular gyrus, and basal forebrain were associated with moral judgment (Moll et al., 2003; Young & Dungan, 2012).

However, it is worth noting that while a domain-specific approach may offer valuable insights into the neural correlates of moral cognition (Barrett, 2012; Binney & Ramsey, 2020), it operates under the assumption that the researchers have exclusively manipulated the moral content in the stimuli. This means that the moral and non-moral stimuli should only vary regarding factors that elicit moral cognition, while other variables are kept constant. Despite efforts to control for confounding differences, studies using the domain-specific approach to compare moral versus non-moral stimuli have struggled to pinpoint brain regions uniquely associated with morality. Instead, they have frequently highlighted neural evidence

of heightened emotional and social processing in the context of moral cognition (Young & Dungan, 2012). Furthermore, the mounting body of evidence suggests that moral cognitive processing involves a wide array of brain regions, thus lacking a clear consensus on which areas are specifically linked to morality (Eres et al., 2018).

2. Domain-General Approach

The transition from a domain-specific approach to a more expansive domain-general perspective marks a pivotal shift in the field of moral neuroscience. This paradigm shift not only underscores the intricacies and interconnectedness of moral cognition within the human brain but also abandons the presumption of a singular, dedicated brain region exclusively devoted to moral processing. Instead, the domain-general approach recognizes that various brain regions and networks collaborate to shape moral judgments, often with their roles overlapping and intertwining. In contrast to the domain-specific approach, the domain-general perspective delves into how diverse cognitive components, previously considered confounding variables, contribute to moral cognition.

This perspective finds consistent support in comprehensive reviews of the neuroscience of morality. Greene and Haidt's (2002) review on moral cognition identified a constellation of brain regions associated with moral processing, including the medial frontal gyrus, posterior cingulate cortex (PCC), precuneus, retrosplenial cortex, orbitofrontal cortex (OFC), amygdala, parietal lobe, dorsolateral prefrontal cortex (dlPFC), superior temporal sulcus (STS), inferior parietal lobe (IPL), and the temporal pole. This observation was largely confirmed in a meta-analysis by Eres et al. (2018). This study compared 84 separate fMRI investigations involving moral judgment tasks with other cognitive tasks, such as legal or

social judgment tasks, revealing multiple brain regions associated with moral cognition. These regions include the medial prefrontal cortex (mPFC), left and right temporoparietal junction (TPJ), left amygdala, precuneus, left inferior orbitofrontal cortex (IOFC), and the insular. This empirical evidence strongly suggests that moral judgment emerges from the orchestrated activity of domain-general cognitive capacities, encompassing perspective-taking, salience processing, executive control, valuation, adherence to social norms, and social decision-making (Yoder & Decety, 2018). Furthermore, the perception of harm, inference of intentionality, emotional processing, empathetic regulation, and theory of mind all contribute to the intricate processes involved in moral cognition (Young & Dungan, 2012).

3. Modularity in Morality

Despite the differences in conceptualizing moral domains, MFT and MAC agree that morality is multifaceted and modular. MFT employs a mechanical metaphor, referring to moral domains as "modules" (Haidt & Joseph, 2004), while MAC uses a chemistry metaphor, terming them "elements" that can be combined into "molecules" (Curry et al., 2022). Despite the variance in the foundational principles of each theory, they both argue that morality is modular, with each domain of morality being capable of dissociation from the others. In this section, we will delve into the debate surrounding modularity in morality and explore the corresponding neuroscientific evidence.

According to Haidt and Joseph (2004), humans possess multiple moral modules, which are cognitive processing systems that predispose individuals to acquire specific moral concerns. These modules, each corresponding to a moral foundation, are akin to systems with

which people are born, and through cultural experiences, these modules further evolve. Thus, the process of moral cognition can be conceived as the input of witnessing a moral transgression (or a moral behavior) into these moral modules, which in turn produce moral intuitions.

On the other hand, Curry et al. (2022) propose that the seven types of cooperation give rise to the seven types of morality. However, these are elemental moral components that can be combined to create a much larger number of complex moral molecules. Moral elements are considered innate and universal, whereas moral molecules are acquired and culturally relative. For example, the combination of group loyalty and respect can form the idea that one ought to help their group by deferring to superior groups, which is the basis of a tribute. The tribute molecule may elucidate why paying taxes is perceived as a moral act in certain cultures. While different terminologies are employed, both MFT and MAC assert that morality is modular. In this dissertation proposal, we use the term “moral modules” when discussing both theories.

Suhler and Churchland (2011) have voiced substantial criticisms of MFT, including the lack of neuroscientific evidence regarding the existence of these modules. They argue that, given MFT's evolutionary, developmental, and psychological underpinnings, it should be supported by biological evidence. Suhler and Churchland contend that each moral foundation does not translate into domain-specific activation in the brain, as moral judgment engages multiple functions. They assert that the modularity of each foundation cannot be identified at the neurological level. While MAC has not been subject to such criticisms due to its recent introduction, it may also be susceptible to the same critique regarding the lack of neuroscientific evidence for moral modules.

In response to this criticism, MFT proponents argue that moral modules are functional models rather than physical or anatomical entities (Haidt & Joseph, 2011). They assert that expecting a specific brain region, or even a single neuron, to function as a moral module is misguided. Moral foundations are not isolated "spots in the brain," and they cannot be reduced to "one specific physiological signature" (Graham et al., 2013, p. 96). Considering that moral judgment emerges from the complex interplay among multiple neural systems, whose functions are typically not (and perhaps never will be) exclusive to moral judgment (Greene & Haidt, 2002), this dissertation proposal extends Haidt and Joseph's conceptualization of moral modularity. It suggests that each module manifests as distinct functional *patterns* across multiple neural networks recruited by various neurons.

This operationalization offers a way to address the ongoing debate about the location of morality in the brain. According to Young and Dungan (2012), moral cognition can be "everywhere," as it relies on complex cognitive capacities and occupies significant space in the brain, or "nowhere," as there are no brain regions uniquely responsible for moral cognition. Our approach seeks to clarify that these contrasting statements are inadequate descriptions of moral cognition. The proposed operationalization shifts the focus from asking where morality resides in the brain to what the unique functional pattern associated with different domains of moral cognition is. This perspective provides a more nuanced and comprehensive understanding of the neural basis of morality, emphasizing the complex interplay of neural networks and cognitive processes involved in moral judgment.

Operationalizing moral modules as distinctive functional patterns across an extensive range of neural networks aligns with existing literature. Wasserman et al. (2017) conducted an analysis of neural representations for harm and purity violations, revealing convergence in

neural activation within a social-cognition or default-mode network encompassing regions such as TPJ, temporal lobes, precuneus, and vmPFC across both moral domains. These domains exhibited different activation patterns in certain regions, with harm-based violations implicating the precuneus and purity-based violations impacting the inferior frontal gyrus (IFG). Recent research by Hopp et al. (2023) similarly identified shared and distinct neural systems underlying moral judgment in the context of the five MFT foundations. Commonly activated regions for the foundations included dmPFC, PCC, precuneus, TPJ, SMA, and V1. Additionally, distinct neural systems showed distributed multivoxel patterns throughout the brain and encompassed regions previously associated with moral processing, especially TPJ, precuneus, and dmPFC. This suggests that commonly activated voxels during moral judgment not only consistently engage but also demonstrate different patterns depending on the moral foundations. Similar findings were observed in Khoudary et al.'s study (2022), employing spatiotemporal partial least squares correlation analysis, a multivariate technique. This study also yielded evidence supporting the existence of moral pluralism and neural modularity, enabling differentiation between social norms and moral foundation violations.

4. Individual Differences in Moral Domain Sensitivity

Moral intuitions, defined as fleeting signals of approval, disapproval, or other emotional responses triggered when discerning patterns in the social landscape (Haidt & Joseph, 2008), serve as the bedrock of our moral judgments, beliefs, and ensuing actions. While the two discussed theoretical frameworks, MFT and MAC, diverge on various fronts, they converge on a central tenet. There are disparities in moral domain sensitivity across diverse cultures and individuals (Atari et al., 2023). To encompass these concepts from both

theories, we adopt the term "moral domain sensitivity" in lieu of "moral intuition salience," commonly employed in MFT research.

Although a comprehensive study examining individual sensitivity to moral domains within MFT and MAC remains elusive, scholars have explored the neural underpinnings of moral judgment under other facets of individual differences. For instance, individuals exhibiting lower moral competence, reflecting their ability to meld emotional and cognitive processes into moral judgment, displayed heightened activation in the left vmPFC and left posterior STS (pSTS) in comparison to those with higher moral competence (Prehn & Heekeren, 2009). Furthermore, individuals showcasing elevated sensitivity to justice motivation, an encompassing construct involving the moral domains of equity and proportionality from MFT and fairness from MAC, exhibited increased activity in the dlPFC, and the functional connectivity between dlPFC and pSTS as well as TPJ (Yoder & Decety, 2018). Hopp (2021) delved into the examination of moral domain sensitivity across all five MFT foundations, revealing that individuals with similarities in their overall moral domain sensitivity, as opposed to foundation-specific similarities, correlated with resemblances in neural responses during moral judgment, notably observed in the dmPFC.

The modulation of moral domain sensitivity in moral judgments holds significance for two key reasons. Firstly, numerous studies investigating the neural correlates of moral judgments have relied on group-level analyses, effectively treating individual differences in information processing as noise (Prehn & Heekeren, 2009). This underscores the imperative need for a rigorous scientific inquiry into the modulation of moral intuition salience. Secondly, the ramifications of these individual differences in moral domain sensitivity have been extensively explored across various domains, encompassing variations in argument

evaluation (Youk et al., 2023), political ideology (Graham et al., 2009), emotional reactions to various moral transgressions (Atari, Mostafazadeh Davani, et al., 2020), religiosity (Yi & Tsang, 2020), vaccine hesitancy (Amin et al., 2017), and patterns of language use (Kennedy et al., 2021). However, our understanding of the cognitive processes and neural representations underlying these modulations of moral domains remains an area necessitating further exploration.

D. Hypotheses and Research Questions

The primary goal of the dissertation is to use neuroscientific approaches to compare MFT and MAC. First, we examine the overlap in the functional neural representation of moral cognition across the two theories. This exploration aims to unveil the extent to which various cognitive processes are involved in moral cognition and whether it can be specifically linked to distinct neural networks. Second, we examine the variations in the neural representation of moral cognition for each moral domain. This provides evidence for the degree to which moral pluralism is also evident at a neurofunctional level. Lastly, we also examine how moral domain sensitivity affects moral cognition. This will provide evidence as to whether moral domain sensitivity explains individual variations in moral cognition.

1. Examining Neural Representation of Moral Cognition

To understand the neurological mechanism related to moral cognition, we investigate overlaps in the neural repressentatino of cognitive processes related to moral domains within and across the two theories. The overlapping neural representation also provides evidence for shared neural networks across both theories. Previous research, including a systematic literature review by Greene and Haidt (2002) and a meta-analysis by Eres et al. (2018),

indicates the involvement of networks in PFC, TPJ, amygdala, precuneus, and OFC during moral judgment. This empirical evidence strongly suggests that moral judgment emerges from the coordinated activity of domain-general cognitive capacities. These capacities include perspective-taking, salience processing, executive control, valuation, adherence to social norms, social decision-making, inference of intentionality, emotional processing, empathetic regulation, and theory of mind (Yoder & Decety, 2018; Young & Dungan, 2012). Recently, Hopp et al. (2023) identified the theory of mind as a reliably recruited fundamental core area for MFT-related moral cognition. As most existing studies have predominantly focused on moral cognition related to MFT domains, we contribute to the moral cognition literature by examining the common neural networks involved in processing MAC domains.

H1: Moral cognition related to MFT domains exhibits overlapping neural representations in brain regions associated with the theory of mind.

RQ1: Which brain regions demonstrate overlapping neural representations for moral cognition related to MAC domains?

RQ2: Which brain regions exhibit overlapping neural representations for moral cognition across the two theories?

2. Comparing Neural Representations Across Moral Domains

Both MFT and MAC embrace the concept of moral pluralism, suggesting the existence of multiple moral domains that are dissociable yet interconnected (Curry et al., 2022; Haidt & Joseph, 2004). Recent neuroscientific studies, including those by Hopp et al. (2023) and Wasserman et al. (2017), provide empirical evidence supporting the notion of dissociable moral domains being represented at the functional level of cognition. However,

no neuroscientific approach has been employed to examine the neural representation across multiple theories. Given that clearer classification results from more dissociable neural representations, this dissertation investigates whether the neural representation of each moral domain can be classified within and across theories.

H2: The neural representation of two MFT domains can be classified with accuracy above chance.

H3: The neural representation of two MAC domains can be classified with accuracy above chance.

RQ3: Is the accuracy of classifying the neural representation of an MFT domain and a MAC domain above chance?

While a direct neuroscientific comparison between the two theories has been lacking, theoretical and survey-based evidence suggests that certain MFT domains are closely related to MAC domains. For example, fairness is theoretically identical in both theories; hence, only one set of fairness vignettes are used in this dissertation. Care in MFT and family values in MAC share similarities, as the latter is considered a specific instance of care guided by kinship (Curry, Mullins et al., 2019). Similarly, loyalty in MFT and group values in MAC exhibit similarities, with group value being a specific instance of loyalty, considering how loyalty in MFT encompasses both group-level and country-level (Zakharin & Bates, 2023). Respect in MAC may encompass authority in MFT, where authority is seen as a specific instance of respect involving obedience to legitimate authority and respect for traditions (Graham et al., 2013), while respect in MAC involves displays of submission, humility, deference, and obedience (Curry, Mullins, et al., 2019). This alignment is supported by factor analysis on surveys of individuals' moral domain sensitivity, as proposed by Curry et al.

(2019). Behaviors related to groups, deference, fairness, and the general category of 'care' are morally relevant for both theories, while purity is not considered morally relevant from a cooperative perspective, as proposed in MAC. Given that more similar moral domains may recruit similar neural networks, the neural representation of these domains may exhibit less dissociability in classification. Therefore, the following hypotheses are proposed:

H4: The classification accuracy of neural representation for care in MFT and family values in MAC will be lower than the classification accuracy for these moral domains with the rest of the moral domains.

H5: The classification accuracy of neural representation for loyalty in MFT and group values in MAC will be lower than the classification accuracy for these moral domains with the rest of the moral domains.

H6: The classification accuracy of neural representation for authority in MFT and respect in MAC will be lower than the classification accuracy for these moral domains with the rest of the moral domains.

H7: The classification accuracy of neural representation for purity in MFT and other MAC domains will be higher than the classification accuracy for other MFT domains and MAC domains.

In addition to examining the classification accuracy across different moral domains, a holistic exploration of the relationship of neural representations can be conducted by considering the distribution of neural representations. Therefore, this dissertation investigates how the neural representation across moral domains can be best modeled.

RQ4: Which representation models best predict the pattern of neural activity across moral domains?

3. Modulating Effects of Moral Domain Sensitivity

While both MFT and MAC recognize the impact of moral domain sensitivity on moral cognition, there is a lack of comprehensive exploration supported by neuroscientific evidence. This dissertation seeks to fill this void by formulating research questions that specifically investigate how moral cognition varies based on moral domain sensitivity.

RQ5: Which brain regions exhibit overlapping neural representation for moral cognition that correlates with moral domain sensitivity?

RQ6: How does the classification accuracy of two moral domains differ between individuals with high and low moral domain sensitivity?

RQ7: Do representation models that incorporate moral domain sensitivity enhance the predictability of the pattern of neural activity across moral domains?

III. Methods

A. General Overview

1. Participants and Procedure

Participants for this study were recruited from the University of California, Santa Barbara. They were offered course credit or monetary compensation, with $5 provided for the survey segment and $50 for the brain imaging segment. A total of 1,021 participants completed the survey portion. Among these participants, 31 voluntarily participated in the brain imaging section. All research protocols received the requisite approvals from the University of Santa Barbara Institutional Review Board.

The survey segment was administered online via Qualtrics. This survey included

various measures related to moral domain sensitivity, employing both the Moral Foundations Questionnaire (MFQ-2) and the Morality-as-Cooperation Questionnaire (MAC-Q), which is elaborated below. In addition to the standard survey items, participants who expressed an interest in partaking in the subsequent brain scanning phase also underwent a preliminary safety screening process conducted by the Brain Imaging Center at the University of California, Santa Barbara. Subsequently, a team of trained research assistants conducted comprehensive phone screenings to assess participants' eligibility for brain imaging. Approximately two weeks after completing the survey, the participants were invited to the Brain Imaging Center.

The brain imaging segment included participants reading a series of vignettes describing moral transgressions. They were presented with a total of 72 vignettes, each requiring a rating of moral wrongness on a scale ranging from 1 (not morally wrong) to 4 (extremely morally wrong). Further details concerning the specific stimuli are discussed below. After brain scanning, participants were again presented with the vignettes outside the scanner. They evaluated the severity, relatability, and believability of each moral vignette.

2. Survey Items

Before the brain imaging segment, participants completed MFQ-2 (Atari et al., 2023) and MAC-Q (Curry, Jones Chesters, et al., 2019). The MFQ-2 features a comprehensive set of 36 items, with six items corresponding to each of the six foundational moral domains (i.e., care, equality, proportionality, loyalty, authority, and sanctity), detailed in Appendix A (all appendices are available in OSF: https://osf.io/enmp8). Participants assessed the extent to which each statement aligned with their self-perception, using a 5-point scale ranging from 1

(slightly describes me) to 5 (describes me extremely well). Composite scores were generated for each foundation by averaging the item scores within that specific domain.

The MAC-Q is divided into two sections, each serving a distinct purpose. It encompasses a total of 42 questions (see Appendix B). In the first section, participants evaluated the moral relevance of all seven domains. In the second section, they expressed their agreement or disagreement with a series of moral judgments presented. Participants rated each item on the MAC-Q using a scale ranging from 0 to 100. Composite scores were computed using three approaches: averaging the scores from the moral relevance section for each domain, averaging the scores from the moral judgment section for each domain, and averaging both the moral relevance and judgment sections for each domain.

A single item was used to measure each moral vignette's severity, relatability, and believability to reduce participant fatigue. Participants viewed each vignette outside of the scanner and indicated their agreement or disagreement (on a 5-point Likert scale) with the following statements: the described action has severe consequences, the described action is related to their personal experience, and the described action is something likely to happen in real life.

3. Stimuli

A total of 72 vignettes were presented to the participants in this study. Among them, 42 vignettes were drawn from the Moral Foundations Vignettes (MFV; Clifford et al., 2015), which provide concise one-sentence descriptions of moral violations related to five fundamental categories (i.e., care, fairness, loyalty, authority, sanctity) and a non-moral transgression (i.e., social norms; see Appendix C for details).

The remaining vignettes were created by the author and an experienced moral scholar to depict violations of MAC domains (see Appendix D). Considering the overlap between fairness described in MFV and MAC, five vignettes were specifically created for each of the six MAC domains: family, reciprocity, bravery, property, respect, and group values. The construction of these vignettes followed the guidelines outlined by Clifford et al.'s (2015) guidelines: avoiding overtly political content, refraining from scenarios reliant on cultural or temporal context, ensuring brevity, readability, and the absence of references to other moral domains.

In the experimental design, these vignettes were arranged following an event-related structure, distributed randomly across three functional runs, and each run lasted approximately six minutes. Participants viewed one vignette at a time and were instructed to immerse themselves in the depicted situation. While each vignette was displayed on the screen for 8 seconds, participants provided a judgment regarding the moral wrongness of the actions described, with a rating scale ranging from 1 (not morally wrong) to 4 (extremely morally wrong). The inter-trial interval (ITI) averaged 4 seconds, with a jitter of 2.16 seconds, to maximize the variability and unpredictability of stimulus presentation.

B. MRI Acquisition and Preprocessing

1. MRI acquisition

The data were acquired using a Siemens Magnetom Prisma 3T MRI system, housed at the Brain Imaging Center at the University of California, Santa Barbara. We acquired all images with the Siemens 64 channel head/neck coil with all elements enabled. We acquired both T1- and T2-weighted anatomical scans using a magnetization-prepared rapid gradient

echo (MP-RAGE) sequence and Turbo Spin-Echo sequences (both 3D) with 0.94 mm isotropic voxels, acquisition matrix of 246 × 256, flip angle (FA) of 7° (T1w) and 120° (T2w), inverse time (TI) of 851ms (T1w), and echo time/repetition time (TE/TR) of 2.22/2500ms (T1w) and 566/3200ms (T2w). We collected 256 slices for both anatomical scans. A field map was acquired with a double-echo spoiled gradient echo sequence with 2mm isotropic voxels, acquisition matrix of 104 × 104, FA of 60°, TE/TR of 7.38/758ms. For all functional scans, multiband (MB) 2D GE-EPI scanning sequence with an MB factor of 8, acquiring 72 2mm interleaved slices with .2mm spacing between slices, isotropic voxel size 2 mm, acquisition matrix of 104 × 104, FA of 52°, TE/TR: 37/720 ms, and anterior-posterior phase-encoded direction to measure blood-oxygen-level-dependent (BOLD) contrast images.

2. Preprocessing

For each of the three functional runs found for 28 subjects (three subjects were excluded for technical errors during data acquisition), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep (Esteban et al., 2020). 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, Jenkinson et al., 2002). The BOLD time series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head motion. These resampled BOLD time series will be referred to as preprocessed BOLD in the original space or just preprocessed BOLD. The BOLD reference was then co-registered to the T1w reference using mri_coreg (FreeSurfer) followed by flirt

(FSL, Jenkinson & Smith, 2001) with the boundary-based registration (Greve & Fischl, 2009) cost-function. Co-registration was configured with twelve degrees of freedom to account for distortions remaining in the BOLD reference. Several confounding time series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS, and three region-wise global signals. FD was computed using two formulations following Power et al. (2014) and Jenkinson et al. (2002). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors was extracted to allow for component-based noise correction (CompCor, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM, and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by thresholding the corresponding partial volume map at 0.05, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are

sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance time series are calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by (Patriat et al., 2017). The BOLD time series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym 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 the 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).

C. Analyses

1. Power Analysis

To calculate the optimal sample size required for detecting significant effects, we conducted a power analysis. It is crucial to note that performing power calculations for fMRI

data introduces complexity, primarily influenced by factors such as the number and duration of runs, alongside the variance of the effect size at every voxel (Hayasaka et al., 2007; Mumford & Nichols, 2008). However, due to the intricacies involved in such calculations, this dissertation followed previous literature and employed an ROI-based power analysis (e.g., Guo et al., 2024). The power analysis was conducted using Gpower 3.1. The chosen parameters included an F-test (repeated measures, within-factors ANOVA), an effect size of 0.25 (i.e., medium effect size), the α error probability set at .05, and the desired power of .8, with seven groups and three measures.

2. Stimuli Validation

In this dissertation, we conducted two validations to ensure the reliability and alignment of our crafted MAC domain vignettes with MFV from Clifford et al. (2015). The first validation focused on comparing participants' evaluations, including moral ratings, believability, relatability, and severity, across social norm vignettes, MFV (encompassing care, loyalty, authority, sanctity, and fairness), and MAC vignettes (covering family, group, reciprocity, heroism, deference, and property). Each evaluation category underwent a one-way Analysis of Variance (ANOVA) with social norm vignettes, MFV, and MAC vignettes as the grouping variable.

We employed advanced large-language models (LLMs) for the second validation, specifically leveraging GPT-3.5, which has demonstrated alignment with human moral judgments (Dillon et al., 2023). Following the presentation of explicit definitions for each moral domain from Graham et al. (2013) and Curry, Mullins et al. (2019), GPT-3.5 was utilized to categorize each vignette into its respective moral domain. The classification

accuracy was then provided for both MFV and MAC vignettes. These validations ensured the robustness and validity of our experimental materials, paving the way for subsequent analyses.

3. Analyses of Moral Domain Sensitivity (Survey)

The means and standard deviations were computed across the participants who completed the survey. To examine the relationships between their moral domain sensitivity, bivariate Pearson correlations were conducted, along with exploratory factor analyses (EFAs). Two EFAs were performed, considering the different measures in MAC-Q. The first EFA included MFT domain sensitivity as well as moral relevance and judgment sections of MAC domain sensitivity. The second EFA included MFT domain sensitivity and averaged MAC domain sensitivity. Bartlett's test of sphericity was conducted to compare the correlation matrix to an identity matrix. A significant value indicated that a factor analysis might be useful. Additionally, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) was assessed to examine the proportion of variance in variables that underlying factors might cause. High values (close to 1.0) generally indicated the adequacy of EFA. EFA was conducted with a varimax rotation, and the factor loadings were calculated for components with eigenvectors above 1.

To examine the representativeness of the MRI participants, two-sample *t*-tests were conducted for each of the moral domain sensitivity measures, comparing the mean moral domain sensitivity for the survey participants and the MRI participants. Additionally, bivariate Pearson correlations were conducted. The means and standard deviations for the MRI participants' moral domain sensitivity were also calculated.

4. Identifying Overlapping Neural Representations of Moral Cognition

To elucidate overlapping neural representations of moral cognition, this dissertation employed an encoding logic utilizing a set of features, specifically moral domains, to explicate neural activities across the brain. In essence, the activity of each voxel was modeled as a linear combination of distinct features represented by moral domains. Parametric statistical tests, specifically repeated-measures *t*-tests, were employed for all comparisons.

For the first-level General Linear Model (GLM), standard preprocessing steps were undertaken (as mentioned above), which provided confounds that are added to the GLM: standard motion correction, three global signals correction (CSF, white matter, and whole-brain), and six anatomical component-based noise corrections estimation for each run. Spatial smoothing was performed using a 6-mm full-width at half-maximum (FWHM) kernel, and highpass temporal filtering was applied via Gaussian-weighted least-squares straight line fitting (sigma = 100 s).

Each model featured an explanatory variable (EV) for each type of vignette (e.g., social norm, care, loyalty, authority, sanctity, fairness, family, group, reciprocity, heroism, deference, and property), convolved with a hemodynamic response function (gamma convolution = 6 s, SD = 3). Planned contrasts were employed to model neural activations unique to each moral domain relative to the social norms condition. First-level models were then advanced into a second-level mixed-effects analysis, pooling across runs within each participant, and subsequently into a group-level mixed-effects analysis, pooling across participants.

Three conjunction analyses were performed separately for MFT domains, MAC domains, and all domains combined to identify overlapping neural representations.

Conjunction analyses were executed using the product of the group-level GLM with a false discovery rate (FDR) correction set at $q < .05$. Following the calculation of the geometric average of the products, a cluster-based thresholding procedure was applied, utilizing a cluster-defining threshold of $Z = 1.96$ and a cluster extent threshold of $p < .05$. The inferences regarding significant clusters were made using meta-analysis from Neurosynth (http://www.neurosynth.org).

5. Comparing Neural Representations Across Moral Domains

As encoding approaches may overlook the nuanced distribution of neural activity, which is critical for understanding how distinct moral domains are represented in the brain (Diedrichsen et al., 2018), this dissertation also employed a decoding approach to compare the neural representations across moral domains. It should be noted that while decoding is a popular approach for analyzing multivariate brain activity patterns, it has limitations when making inferences. Just because we can successfully decode feature X (moral domains in this case) from brain region A does not mean that the representation in A is exclusively defined by feature X. There could be many other features influencing activity patterns in that region.

Multivoxel pattern analysis (MVPA) utilizing support vector machines (SVM) was employed to decode the neural representation of moral cognition. SVM, a machine learning algorithm, classified neural representation into a finite set of moral domains, offering insights into how the brain differentially processes moral information. Given computational constraints, MVPA was executed iteratively for all possible pairs of moral domains. First, a pair of moral domains was selected, and beta weights of contrasts from second-level GLM for the two moral domains were pooled. Utilizing 10-fold cross-validation, a subset of beta

weights served as the training dataset, while the remaining served as the testing dataset. The trained SVM was then deployed to predict the moral domain based on the distribution of neural activity in the test data. The average accuracy was calculated for classification within MFT, within MAC, and across MFT and MAC. Additionally, one-sample *t*-tests were conducted to test H4, H5, and H6. For example, to test H4, the accuracy of classifying care in MFT and family value in MAC was compared against the classification accuracies for care with 9 other moral domains and family value with 9 other moral domains. To test H7, an independent samples *t*-test was conducted. The classification accuracies for purity and 6 MAC domains (fairness was excluded) were compared with the classification accuracies for three MFT domains (i.e., care, loyalty, and authority) and six MAC domains.

While the decoding approach provides insights into differences in neural representation, it falls short of revealing how neural representations of moral domains are interrelated. Rather than scrutinizing all pairwise similarities or differences across moral domains, the neural activation can be comprehensively understood by characterizing a composition of patterns between these domains. Consequently, Pattern Component Modeling (PCM) was employed to explore relationships in neural representations across moral domains. PCM predicted the neural representation of various moral cognitions using predefined patterns of relationships between moral domains.

PCM is a Bayesian approach to examining representation models (Diedrichsen et al., 2018). Unlike decoding approaches that emphasize the spatial arrangement of neural activities, PCM focuses on describing the shape of the distribution of neural activity, offering a powerful and flexible tool for comparing various models, which outperforms representational similarity analysis. Similar to encoding models, PCM evaluates the model's

capacity to predict brain activity patterns. However, unlike encoding models, PCM does not directly fit the activity of individual voxels. Instead, it aims to predict the specific structure of the second-moment matrix (covariance matrix) of neural activity, a central statistical quantity determining the representational content of brain activity patterns.

PCM was implemented using beta weights derived from 11 first-level GLM contrasts, which represent different moral domains (i.e. care, loyalty, authority, sanctity, fairness, family, group, reciprocity, heroism, deference, and property). Various representation models between the 11 moral domains were constructed. The null model posits that all moral domains were equally dissociable from each other, represented by an identity matrix for an 11 x 11 covariance matrix. See Appendix E for a matrix representation of the models.

Several fixed models were created, including the MAC model (MAC domains were related to each other, MFT domains were unrelated to each other) and the MFT model (MFT domains were related to each other, MAC domains were unrelated to each other). Additionally, a fixed model was designed to explore the relationships suggested by H4, H5, and H6 (i.e. H model). The H model predicted that care-family, loyalty-group, and authority-respect domains were related, while other domains were equally dissimilar. Two component models were constructed using a linear combination of the MAC and MFT model (i.e., MAC + MFT model) as well as a linear combination of all three fixed models (i.e., MAC + MFT + H model). A free model, derived from the maximum-likelihood estimate of the second-moment matrix in the presence of noise, was also employed to estimate the flexibility of model fitting and provided corrected correlation estimates between different patterns, in this case, moral domains.

The marginal likelihood of the second-moment matrix under each model was calculated, and the relative likelihood of the data was assessed by computing log Bayes factors. The log Bayes factor was standardized, with 0 indicating the null model and 1 representing the noise ceiling derived from the free model. This provided Psuedo-R^2 values for each fixed model. A positive value indicates the estimates of the free model being explained above and beyond the null model by a given model. A negative value indicates that a given model underperforms compared to the null model. A predicted second-moment matrix was calculated using the model with the highest log Bayes factor. The second-moment matrix was converted into a similarity matrix, where a higher number indicates higher similarity in neural representation.

6. Examining Modulation of Moral Domain Sensitivity

To delve into the impact of moral domain sensitivity on the mechanisms of moral cognition, this dissertation integrated moral domain sensitivity into both encoding and decoding analyses. In the encoding approach, a group-level mixed-effects analysis was conducted by pooling data across participants. This involved examining each contrast derived from the second-level mixed-effects analysis, which aggregated runs within each participant. The previously mentioned second-level model was utilized, incorporating participants' survey responses for moral domain sensitivity as covariates in the higher-level GLM.

Four higher-level GLMs were conducted with different sets of covariates. The first set encompassed moral domain sensitivity for MFT, comprising care, equality, proportionality, loyalty, authority, and sanctity. Additionally, three sets of moral domain sensitivity were considered for MAC: moral relevance, moral judgment, and the average of these scores for

the seven MAC domains. The continuous measures of moral domain sensitivity were discretized into ordinal categories, creating four groups corresponding to each quartile of the moral domain sensitivity scores. Contrasts were calculated based on the participant-level characteristics (i.e., moral domain sensitivity). In other words, trend contrasts were calculated for each measure of moral domain sensitivity.

Conjunction analyses were conducted across different types of moral vignettes for each moral domain sensitivity with the same FDR correction and clustering threshold as outlined above. Consequently, these analyses systematically explored the relationship between moral domain sensitivity and neural signals across various moral vignettes, shedding light on the nuanced interplay between moral sensitivity and neural representation in response to diverse moral scenarios.

MVPA using SVM was performed to compare the neural representations across moral domain sensitivity. Participants were divided based on their moral domain sensitivity, with SVM analyses conducted separately for those with high sensitivity and those with low sensitivity in each moral domain. For example, SVM analysis was applied to participants with care domain sensitivity above the median, classifying care against all other moral domains (i.e., loyalty, authority, sanctity, fairness, family, group, reciprocity, heroism, deference, and property). A parallel analysis was carried out for participants with care domain sensitivity below the median. The differences in classification accuracy between the two groups were calculated. A one-sample *t*-test with a test value of 0 was conducted on the differences. This process was repeated for all moral domains. Considering multiple comparisons, Bonferroni correction was implemented (*p*-values below .0045 were considered statistically significant).

To incorporate moral domain sensitivity in PCM, three component models were introduced. These models represent a linear combination of the MAC + MFT + H model, each supplemented with different measures of moral domain sensitivity. The augmentation was achieved by incorporating the MFQ-2 results with either the moral relevance section, the moral judgment section, or the average of the two sections. Subsequently, the log Bayes factor and predicted second-moment matrix was computed, following the previously outlined procedure.

7. Robustness Checks

Given the inherent individual variations in moral ratings, believability, severity, and relatability attributed to each vignette, this dissertation sought to examine the robustness of previously discussed analyses, including conjunction analysis, SVM, PCM, and the moderating effects of moral domain sensitivity. The initial step involved implementing a first-level GLM with parameters outlined previously. Notably, four additional EVs were introduced, accounting for moral ratings, believability, severity, and relatability, with raw values serving as weights. Planned contrasts were conducted with these moral vignette evaluations as covariates to capture the unique neural signatures associated with each moral domain relative to the social norms condition. Subsequently, these refined first-level models were incorporated into higher-order GLMs and other analyses.

In addition, an exploratory analysis was conducted on the free model of PCM. The free model provided a data-driven prediction of the second-moment matrix (i.e., the covariance of neural representation across moral domains). EFA was then conducted on the second-moment matrix to examine the underlying components of neural representations

across moral domains. The EFA parameters and procedure were identical to the EFA on the survey data.

IV. Results

A. Power Analysis

The conducted power analysis indicated that the minimum sample size required for robust statistical power is 28, aligning with the actual number of participants analyzed in this dissertation. This congruence provides evidence that the chosen sample size was sufficient to effectively detect significant effects in the context of the experimental design.

The average age of these 28 participants was 20.28 (SD = 2.05). Seventeen participants identified their biological sex as female. Racial identity distribution included 7 Whites, 7 Latinas, 5 Asians, 2 African Americans, and the remainder had mixed racial backgrounds. The majority of participants indicated English as their first language (n = 24).

B. Stimuli Validation

The ANOVA results predominantly indicated that participants ascribe different levels of moral ratings, believability, relatability, and severity to MFV, MAC vignettes, and social norm vignettes. In terms of moral ratings, MAC vignettes exhibited the highest scores for moral wrongness (M = 3.15, SD = 0.93), followed by MFV (M = 2.98, SD = 0.98), and social norm vignettes (M = 1.35, SD = 0.64), $F(2, 1964) = 302.65$, $p < .001$. Notably, participants perceived moral transgressions in MAC vignettes to be more believable (M = 3.96, SD = 1.08) compared to MFV (M = 3.50, SD = 1.29), followed by social norm vignettes (M = 2.98, SD = 1.32), $F(2, 1929) = 62.84$, $p < .001$. Additionally, MAC vignettes (M = 2.00, SD =

1.29) were considered more relatable than MFV (M = 1.74, SD= 1.17) and social norm vignettes (M = 1.75, SD = 1.18), $F(2, 1929) = 9.88$, $p < .001$. However, there is no significant difference between the relatability of MFV and social norm vignettes. MAC vignettes receive the highest severity ratings (M = 4.09, SD = 1.06), followed by MFT (M = 3.65, SD = 1.20), and social norm vignettes (M = 1.46, SD = 0.84), $F(2, 1929) = 437.78$, $p < .001$. Table 1 provides the means and standard deviations for moral ratings, perceived believability, relatability, and severity of the vignettes across each moral domain (refer to Figure 1 for visualization). Pairwise comparisons are detailed in Appendix E.

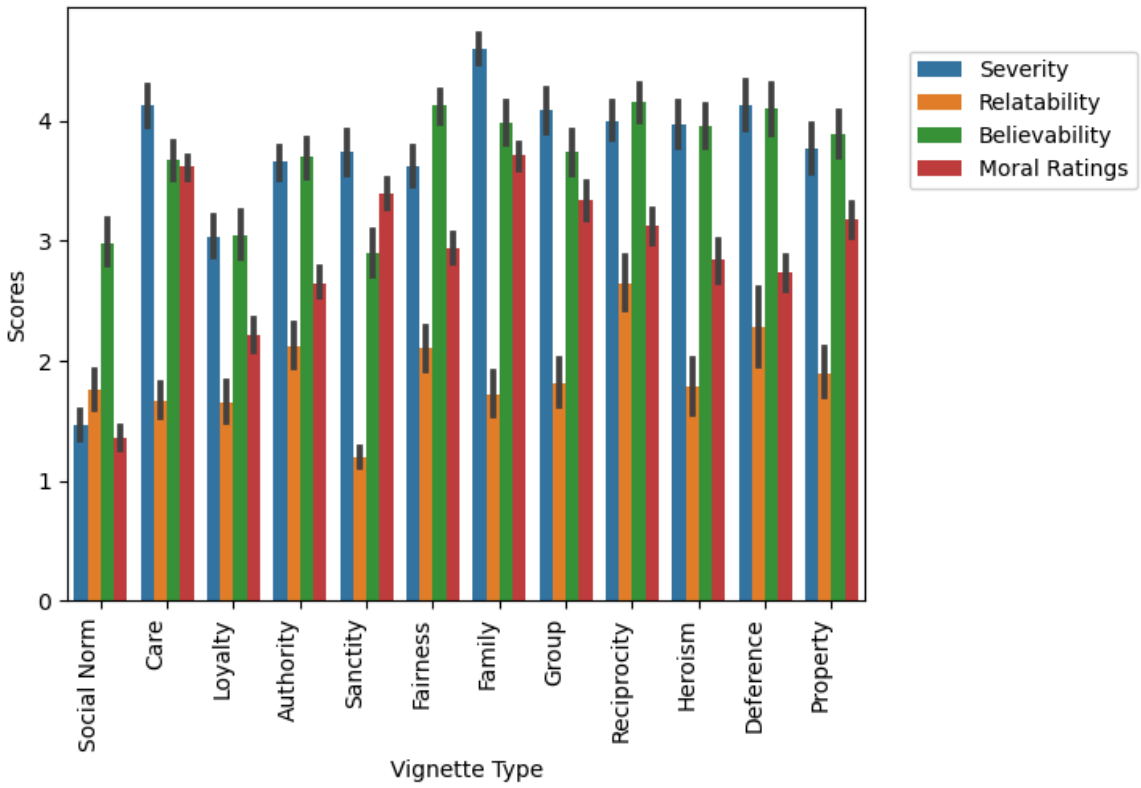
Table 1

Means and Standard Deviations of Moral Vignette Evaluations

	Believability		**Relatability**		**Severity**		**Moral ratings**	
Vignette Type	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Social Norm	2.98	1.32	1.76	1.18	1.46	0.84	1.36	0.64
Care	3.67	1.19	1.67	1.06	4.12	1.14	3.61	0.69
Loyalty	3.05	1.32	1.65	1.11	3.04	1.13	2.22	0.85
Authority	3.70	1.20	2.12	1.34	3.65	0.99	2.65	0.83
Sanctity	2.90	1.34	1.19	0.59	3.74	1.28	3.39	0.92
Fairness	4.12	0.97	2.10	1.31	3.62	1.22	2.93	0.91
Family	3.99	1.07	1.72	1.08	4.60	0.74	3.71	0.64
Group	3.73	1.13	1.81	1.18	4.09	1.08	3.34	0.93
Reciprocity	4.16	0.99	2.64	1.47	4.00	0.97	3.12	0.85
Heroism	3.95	1.05	1.79	1.21	3.96	1.15	2.84	1.05
Deference	4.10	1.00	2.27	1.44	4.13	0.98	2.73	0.82
Property	3.89	1.15	1.89	1.15	3.76	1.20	3.18	0.90

Figure 1

Bar Graph of Moral Vignette Evaluations



Regarding the LLM-based validation, the findings underscored the capability to accurately classify both MFV and MAC vignettes above the chance level of 9%. The overall accuracy in correctly categorizing the vignettes was 47.69% (refer to Table 2). Notably, there was a lower incidence of misclassification within theories as opposed to across theories. Specifically, only 20% of the MFV instances were misclassified as other MFT domains, while 15.38% of the MAC vignettes found classification within other MAC domains. In stark contrast, 56.67% of MAC vignettes were classified within MFT domains, while 26.47% of MFV instances were classified within MAC domains.

Table 2

Crosstab of LLM-based Classification

	Predicted Classification										
Actual Classification	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Care	**2**	0	0	1	1	0	0	0	0	2	0
2. Loyalty	0	**4**	0	0	2	0	1	0	0	0	0
3. Authority	0	0	**1**	1	0	1	0	0	0	4	0
4. Sanctity	0	0	0	**6**	0	1	0	0	0	0	0
5. Fairness	0	0	0	0	**7**	0	0	0	0	0	0
6. Family	2	1	0	0	1	**1**	0	0	0	0	0
7. Group	0	0	0	0	3	0	**2**	0	0	0	0
8. Reciprocity	0	0	0	0	1	0	1	**3**	0	0	0
9. Heroism	3	1	0	0	1	0	0	0	**0**	0	0
10. Deference	0	0	3	0	0	0	0	1	0	**1**	0
11. Property	0	0	0	0	1	0	0	0	0	0	**4**

Although the two validations indicated that MFV and MAC vignettes may be qualitatively different from their theoretical frameworks, they provided evidence for three key points. First, the validations offered compelling evidence regarding the efficacy of MAC vignettes in eliciting moral cognition. The higher perceived believability and relatability of MAC vignettes compared to MFV suggest that participants were more likely to engage in moral cognition when presented with MAC scenarios. This underscored the significance of MAC vignettes as potent tools for probing moral cognitive processes, given participants' heightened ability to envision themselves in the depicted situations.

Second, a robustness check was imperative to address variations in moral ratings and the perceived severity of the moral transgressions outlined in the vignettes. The higher moral ratings and perceived severity for MAC vignettes, followed by MFV and social norm vignettes, emphasized the importance of statistically accounting for these factors. While social norm vignettes served as controls, their portrayal of non-moral and less consequential transgressions indicated potential confounds.

Lastly, the LLM validation revealed that most classification errors occurred across theories, such as classifying MAC vignettes as MFV or vice versa. This observation demonstrated the importance of comparing moral domains within a theory and across theories—a central focus of this dissertation. Examining differences in neural responses to moral domains within and across theories would contribute a valuable dimension to understanding the neural underpinnings of moral pluralism.

C. Analyses of Moral Domain Sensitivity (Survey)

Table 3 presents the means and standard deviations of moral domain sensitivity for participants who completed the survey (n = 1,021), including the 31 participants who participated in the brain imaging section. Overall, the participants demonstrate high sensitivity to the care foundation (M = 4.03, SD = 0.69) and relatively low sensitivity to sanctity/purity (M = 2.30, SD = 0.82) within the MFT domains. Regarding MAC domains, particularly for the averaged score across the moral relevance and judgment section, participants exhibit high sensitivity to fairness (M = 67.60, SD = 14.89) but low sensitivity to deference (M = 43.81, SD = 18.64).

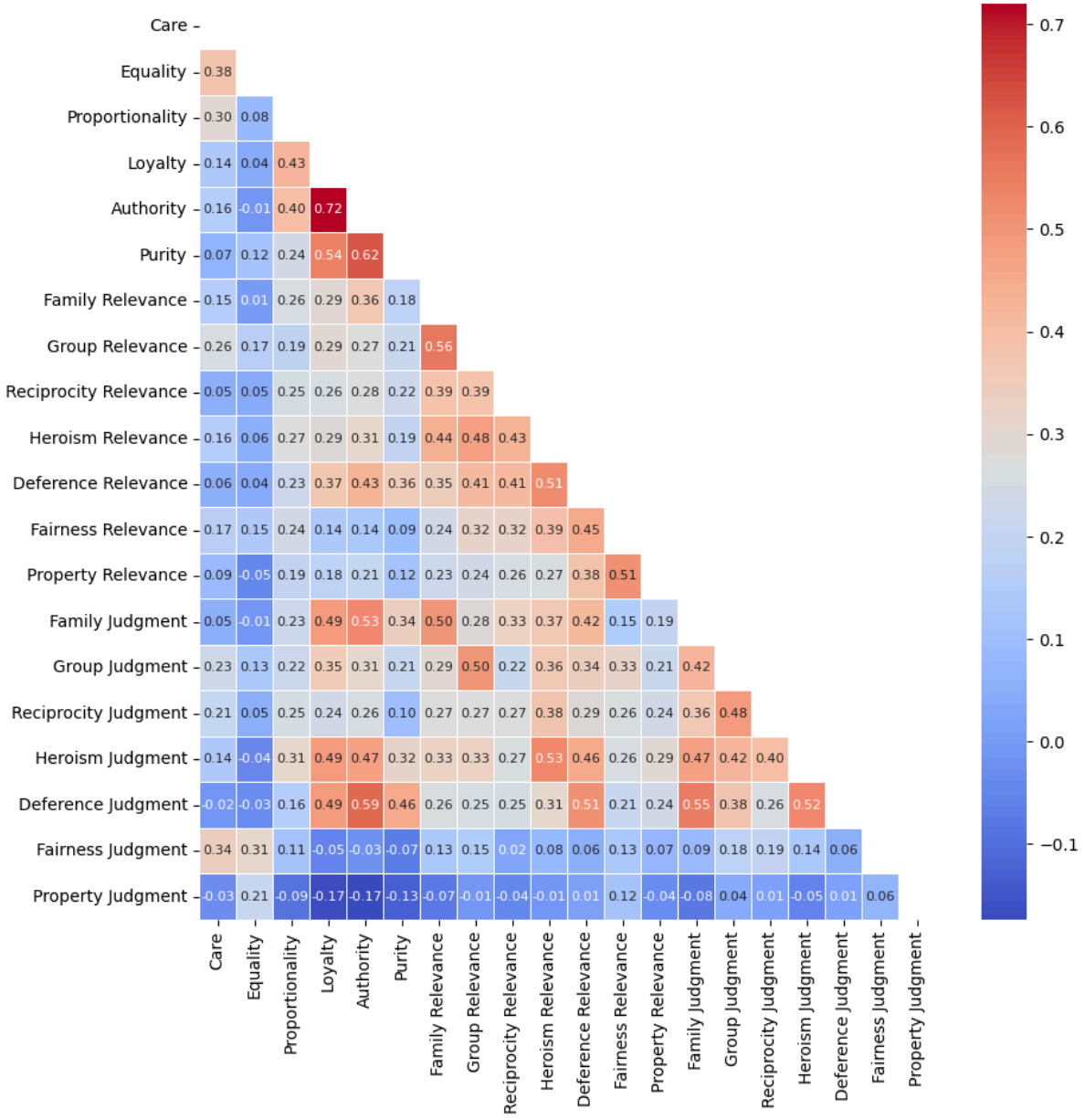
Table 3

Means and Standard Deviations of Moral Domain Sensitivity for Surveyed Participants

	Mean (SD)		
Care	4.03 (0.69)		
Equality	2.73 (0.87)		
Proportionality	3.66 (0.64)		
Loyalty	2.74 (0.79)		
Authority	3.06 (0.80)		
Purity	2.30 (0.82)		
	Relevance	Judgment	Overall
Family	70.96 (20.18)	56.45 (23.55)	63.66 (19.26)
Group	61.46 (20.23)	57.78 (18.62)	59.57 (17.01)
Reciprocity	66.96 (24.53)	71.45 (17.23)	69.14 (16.91)
Heroism	58.95 (21.82)	59.84 (19.09)	59.35 (18.03)
Deference	46.60 (21.99)	41.14 (20.38)	43.81 (18.64)
Fairness	58.04 (22.00)	77.24 (17.38)	67.60 (14.89)
Property	60.40 (25.00)	42.75 (20.04)	51.59 (16.28)

Figure 2

Bivariate Correlations of Moral Domain Sensitivity for Surveyed Participants



As there were three different measures of moral domain sensitivity for MAC, two EFAs were conducted. The first EFA encompassed MFT domain sensitivity, along with the moral relevance and judgment sections of MAC domain sensitivity. Bartlett's test of sphericity ($\square^2 = 7651.22$, $p < .001$) and the KMO test (KMO = .86) affirmed the adequacy of

conducting EFA. The results revealed that a three-factor model explained 42.85% of the variance in the moral domains (refer to EFA 1 in Table 4 for factor loadings). The eigenvalues were close to 1 for all subsequent factors.

The second EFA incorporated MFT domain sensitivity and averaged MAC domain sensitivity across relevance and judgment sections. Bartlett's test of sphericity ($\chi^2 = 4947.45$, $p < .001$) and the KMO test (KMO = .85) justified the suitability for conducting EFA. The results demonstrated that a three-factor model explained 50.00% of the variance in the moral domains (see EFA 2 in Table 4 for factor loadings). The eigenvalues were close to 1 for all subsequent factors.

The factor loadings from both EFAs collectively suggested that care and equality loaded on one factor, while loyalty, authority, and purity loaded on another. In contrast, MAC domains were predominantly loaded on a separate factor. This indicated that different underlying constructs influence the measures of MFT domain sensitivity and MAC domain sensitivity.

Table 4

Factor Loadings for EFA

	EFA 1			EFA 2		
	F1	F2	F3	F1	F2	F3
Care	0.10	0.07	**0.75**	0.11	0.11	**0.71**
Equality	0.01	-0.03	**0.53**	-0.01	0.06	**0.53**
Proportionality	0.26	0.32	0.27	0.39	0.21	0.27
Loyalty	0.26	**0.78**	0.08	**0.81**	0.16	0.07
Authority	0.27	**0.86**	0.05	**0.88**	0.18	0.02
Purity	0.15	**0.67**	0.02	**0.68**	0.06	0.04
Family (R)	**0.57**	0.19	0.11	0.46	**0.54**	-0.03
Family (J)	0.48	**0.49**	0.00	-	-	-
Group (R)	**0.61**	0.11	0.25	0.29	**0.60**	0.22
Group (J)	**0.52**	0.23	0.28	-	-	-
Reciprocity (R)	**0.55**	0.13	0.01	0.26	**0.62**	0.08
Reciprocity (J)	**0.49**	0.13	0.22	-	-	-
Heroism (R)	**0.69**	0.14	0.08	0.40	**0.64**	0.03
Heroism (J)	**0.54**	0.43	0.07	-	-	-
Deference (R)	**0.66**	0.30	-0.03	0.53	**0.59**	-0.09
Deference (J)	0.44	**0.56**	-0.08	-	-	-
Fairness (R)	**0.59**	-0.05	0.16	-0.05	**0.60**	0.43
Fairness (J)	0.16	-0.10	**0.48**	-	-	-
Property (R)	**0.49**	0.06	0.02	-0.05	**0.45**	0.10
Property (J)	0.06	-0.20	0.08	-	-	-

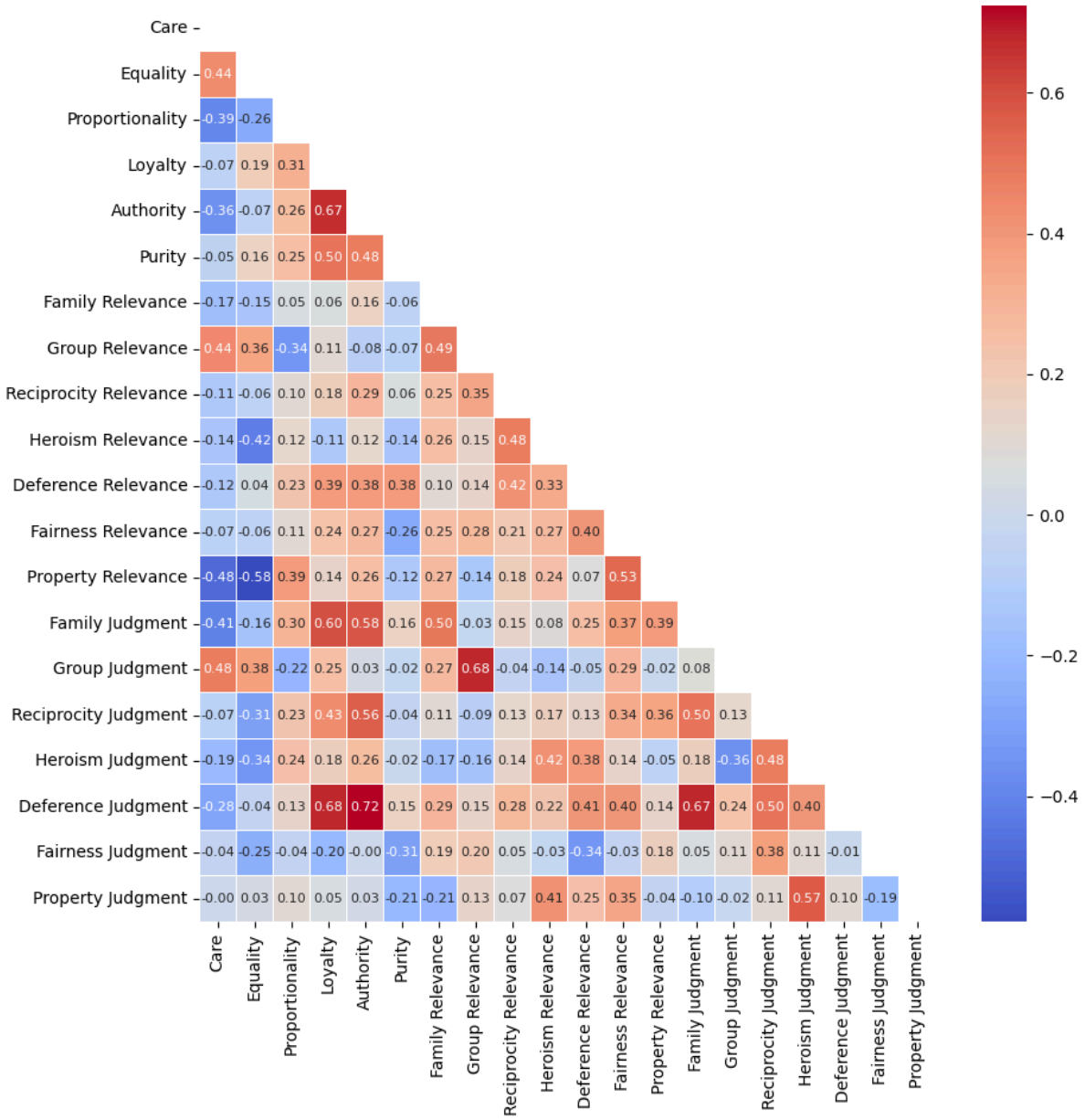
The representativeness of the 28 participants included in the brain analysis was assessed through two-sample *t*-tests and a correlation analysis. Each moral domain sensitivity underwent a two-sample *t*-test, revealing no statistically significant differences across all moral domain sensitivities. The means and standard deviations of moral domain sensitivity for participants in the brain analysis are presented in Table 5.

Table 5

Means and Standard Deviation of Moral Domain Sensitivity for MRI Participant

	Mean (SD)		
Care	4.17 (0.60)		
Equality	2.74 (0.82)		
Proportionality	3.62 (0.53)		
Loyalty	2.52 (0.64)		
Authority	3.15 (0.71)		
Purity	2.27 (0.84)		
	Relevance	Judgment	Overall
Family	76.58 (15.09)	51.40 (23.06)	63.99 (16.63)
Group	59.99 (22.82)	53.01 (19.28)	56.50 (19.28)
Reciprocity	66.02 (28.61)	68.27 (19.44)	67.15 (18.29)
Heroism	60.61 (22.12)	57.50 (19.73)	59.05 (17.66)
Deference	44.90 (23.29)	39.45 (17.83)	42.56 (17.75)
Fairness	58.81 (23.44)	76.96 (12.53)	67.89 (13.10)
Property	61.32 (25.86)	43.18 (13.99)	52.25 (14.47)

Figure 3
Bivariate Correlation of Moral Domain Sensitivity for MRI Participants



D. Identifying Overlapping Neural Representations of Moral Cognition

A conjunction analysis focusing on MFT domains revealed the existence of several shared neural networks, as illustrated in Figure 4 and detailed in Table 6, outlining significant clusters. Noteworthy activations were observed in regions, including the precuneus, PCC,

and TPJ—implicated in the theory of mind processes. Additionally, the superior temporal gyrus associated with sentence comprehension and the angular gyrus associated with semantic processing exhibited notable activations. Therefore, the data is consistent with H1. Moral cognition related to MFT domains exhibited overlapping neural representations in brain regions associated with the theory of mind.

In contrast, a conjunction analysis on MAC domains revealed significant clusters in the precuneus and V1 (see Figure 5 and Table 7). When comparing across all 11 moral domains, the precuneus emerges as a significant cluster with MNI coordinates of 0, -65, 42, a peak activation of 5.44, and a cluster size of 292 mm^3. To answer RQ1 and RQ2, the precuneus, which is associated with the theory of mind, and the primary visual cortex showed overlapping representations for moral cognition related to MAC. In contrast, only precuneus was associated with moral cognition across the two theories.

Table 6

Results of Conjunction Analysis on MFT Domains

Cluster ID	MNI Coordinates			Peak Activation (Z)	Cluster Size (mm^3)
	X	Y	Z		
1	-0.12	-64.25	42.45	5.62	1598
1a	2.70	-60.48	34.92	5.20	
1b	-0.12	-53.90	28.33	4.45	
1c	-10.46	-58.60	38.68	4.18	
2	-59.34	-7.79	-13.07	4.73	483
2a	-51.82	-7.79	-15.89	4.32	
3	58.16	-63.31	18.92	4.47	112
4	-49.94	-70.84	18.92	4.26	273
4a	-56.52	-65.19	17.04	3.84	
5	50.64	-53.90	20.80	4.11	39
6	-55.58	-68.01	19.86	3.71	19.00

Figure 4

Glass Brain Plotting of Conjunction Analysis on MFT Domains

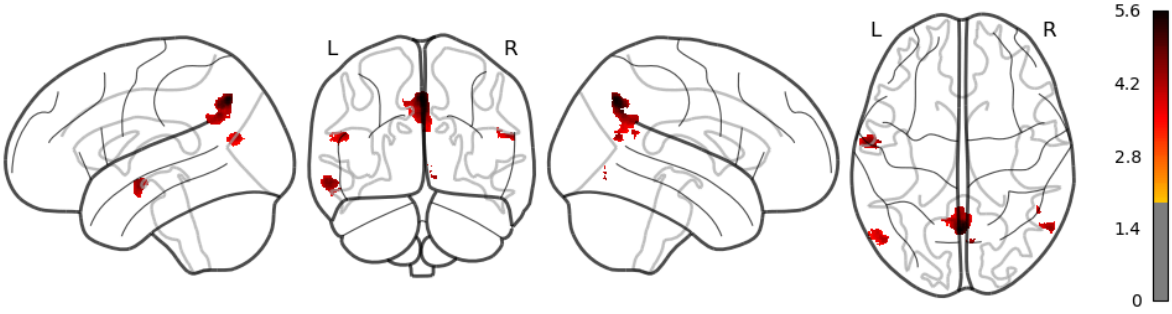


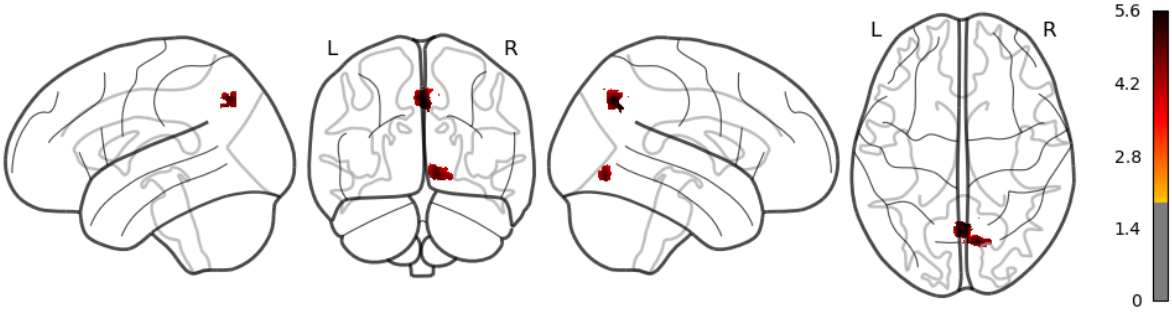
Table 5

Results of Conjunction Analysis on MAC Domains

Cluster ID	MNI Coordinates			Peak Activation (Z)	Cluster Size (mm^3)
	X	Y	Z		
1	1.76	-61.43	37.74	5.64	684
2	10.22	-71.78	-7.42	5.23	465

Figure 7

Glass Brain Plotting of Conjunction Analysis on MAC Domains



E. Comparing Neural Representations Across Moral Domains

The classification accuracy for every combination of moral domain pairs is presented in Figure 8. The overall accuracy was 83.25% (SD = 18.13%). Specifically, the accuracy for classifying moral domains within MFT was 87.94% (SD = 14.16%); within MAC, it was 79.37% (SD = 20.27%). These findings offer compelling evidence that neural representations of moral domains, as conceptualized by MFT and MAC, can be distinctly identified. Hence, the data is consistent with H2 and H3. The classification accuracy across theories (i.e., classifying one domain from MFT and the other from MAC was 84.70% (SD = 16.90%),

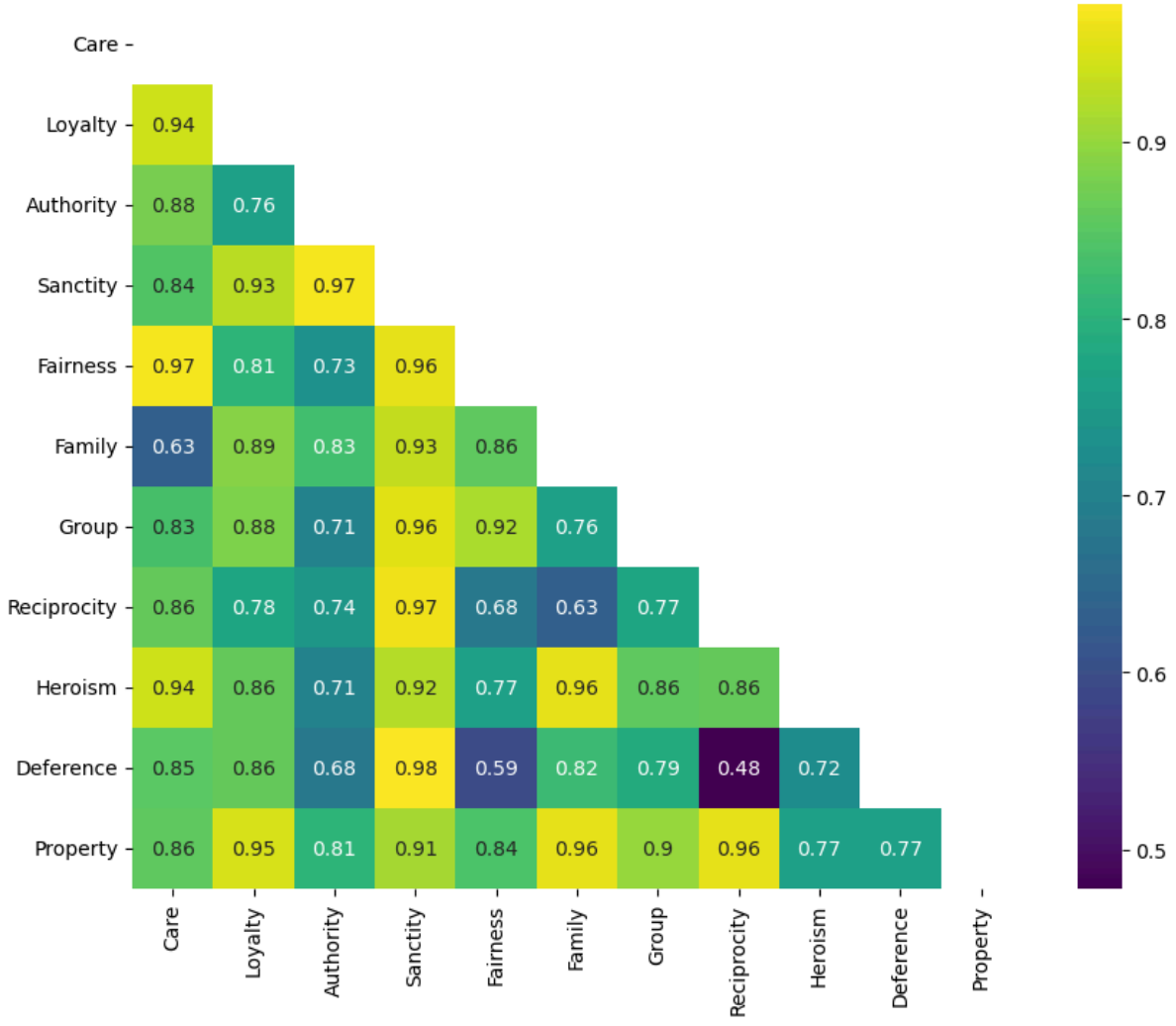
which answers RQ3. The accuracy of classifying the neural representations of moral domains from two different theories was above chance.

For each moral domain, the average classification accuracy was as follows: care = 86.05% (*SD* = 15.36%), loyalty = 86.55% (*SD* = 13.59%), authority = 78.11% (*SD* = 19.56), sanctity = 93.72% (*SD* = 9.43), fairness = 81.28% (*SD* = 18.17%), family = 82.78% (*SD* = 19.03%), group = 83.72% (*SD* = 17.38%), reciprocity = 77.39% (*SD* = 21.96%), heroism = 83.61% (*SD* = 18.01%), deference = 75.39% (*SD* = 21.16%), and property = 87.17% (*SD* = 15.39%). See Appendix G for averaged SVM coefficients for each moral domain, providing a visualization of the neural representation of moral cognition for each domain.

Consistent with H4, the classification accuracy of neural representation for care in MFT and family value in MAC was lower (63%) compared to the classification accuracy for these moral domains with the rest of the moral domains ($M = 86\%$), $t(17) = 11.09$, $p < .001$. Inconsistent with H5, the classification accuracy of neural representation for loyalty in MFT and group values in MAC was not lower (88%) than the classification accuracy for these moral domains with the rest of the moral domains ($M = 84\%$), $t(17) = 1.73$, $p = .10$. Consistent with H6, the classification accuracy of neural representations for authority in MFT and deference in MAC was lower (68%) compared to the classification accuracy for these moral domains with the rest of the moral domains ($M = 78\%$), $t(17) = 3.44$, $p = .003$. Consistent with H7, the classification accuracy of neural representation for purity in MFT and other MAC domains ($M = 95\%$) was higher than the classification accuracy for other MFT domains and MAC domains ($M = 82\%$), $t(22) = 3.47$, $p = .002$.

Figure 8

Classification Accuracy of SVM for Pairs of Moral Domains



To comprehensively investigate the interrelations among moral domains, PCM was conducted with four fixed models. The findings reveal that the MAC model, where MAC domains were related to each other and MFT domains were unrelated to each other, offered a superior representation of the neural underpinnings of moral cognition compared to the null model, which predicted that all 11 moral domains are unrelated to each other (see Figure 9).

Specifically, the MAC model accounts for 5.9% of the relative log Bayes factor in the free model. Contrastingly, other models that expected similarities within MFT or were based on conceptual similarities (as predicted in H4, H5, and H6) exhibited lower performance compared to the null model. A linear combination of MAC and MFT models accounts for 6.4% of the relative log Bayes factor in the free model. Adding the conceptual similarity model into the linear combination did not change the relative log Bayes factor. To better understand the linear combination of the MAC and MFT model, the predicted second-moment matrix is provided in Figure 10. To answer RQ4, the results of PCM indicate that neural representations within MAC domains were more similar than those within MFT domains. The neural representation across the two theories was less similar than the similarities within theories.

Figure 9

Standardized Log Bayes Factor for PCM

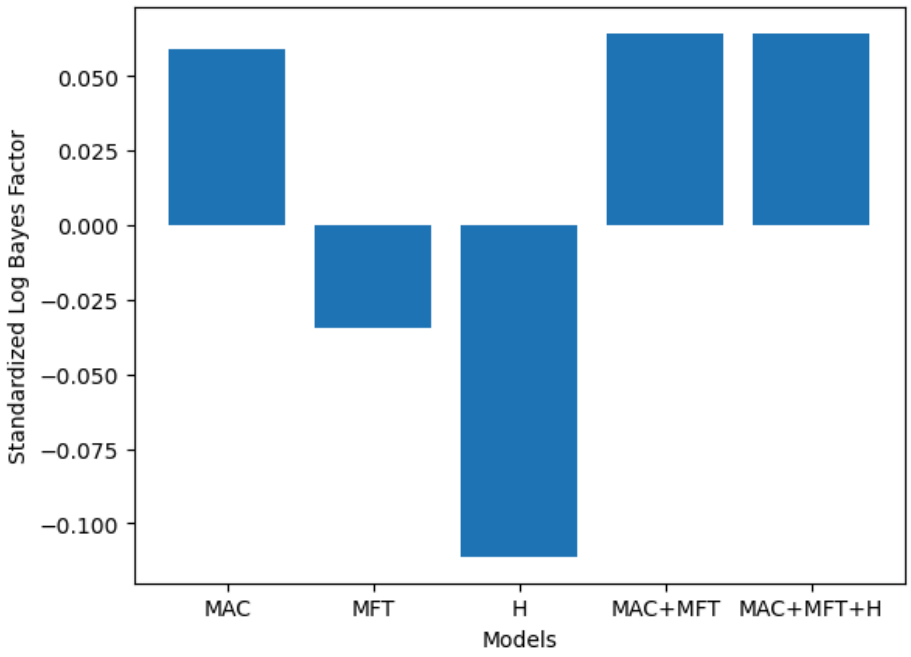
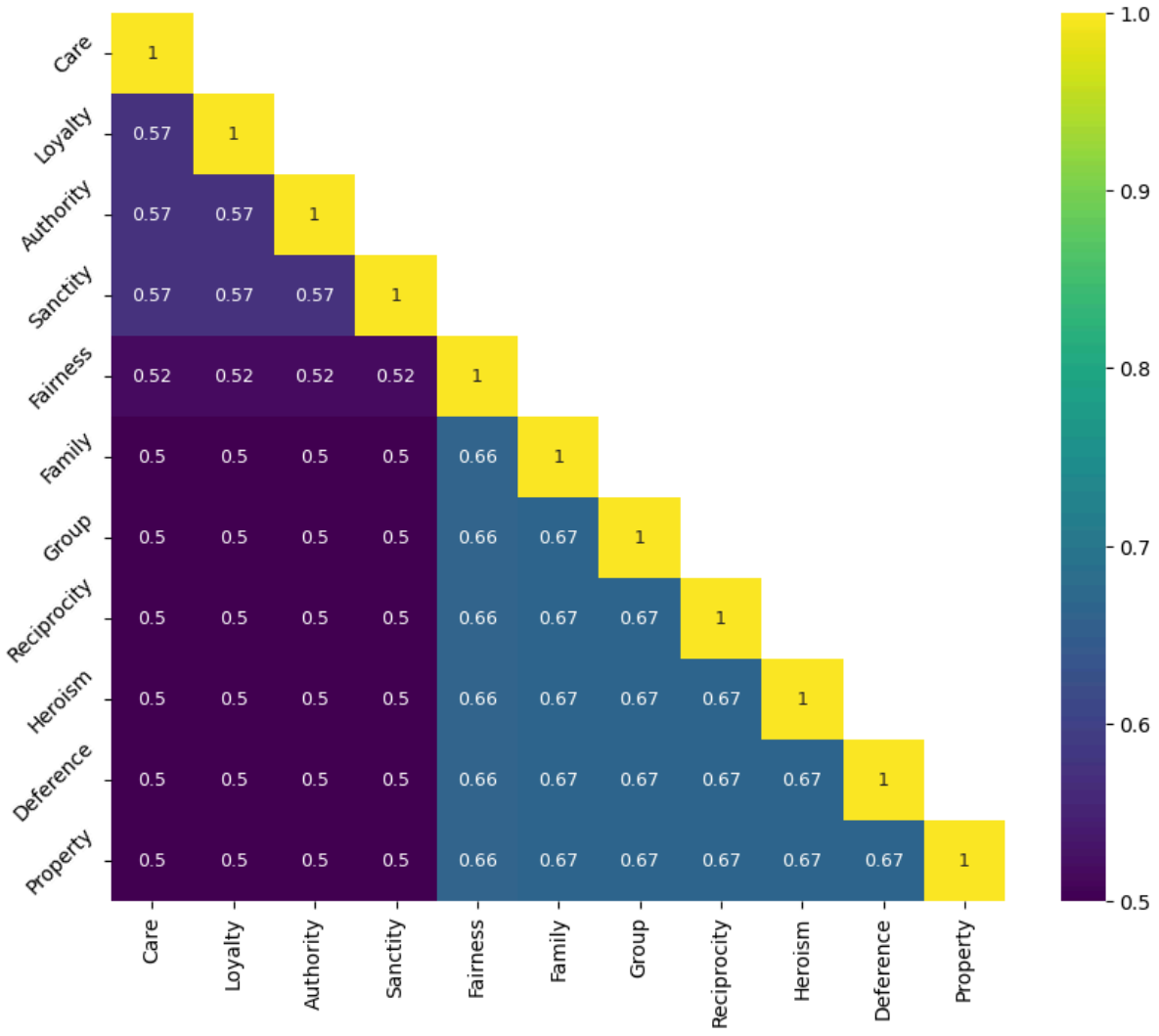


Figure 10

Predicted Similarity Matrix Using MAC + MFT Model



Note. 1 indicates perfect similarity.

F. Examining Modulation of Moral Domain Sensitivity

1. Identifying Overlapping Neural Representations of Moral Cognition

The results of the conjunction analyses yielded the absence of statistically significant findings. To answer RQ5, the results imply the absence of shared neural representation across various moral vignettes related to participants' moral domain sensitivity.

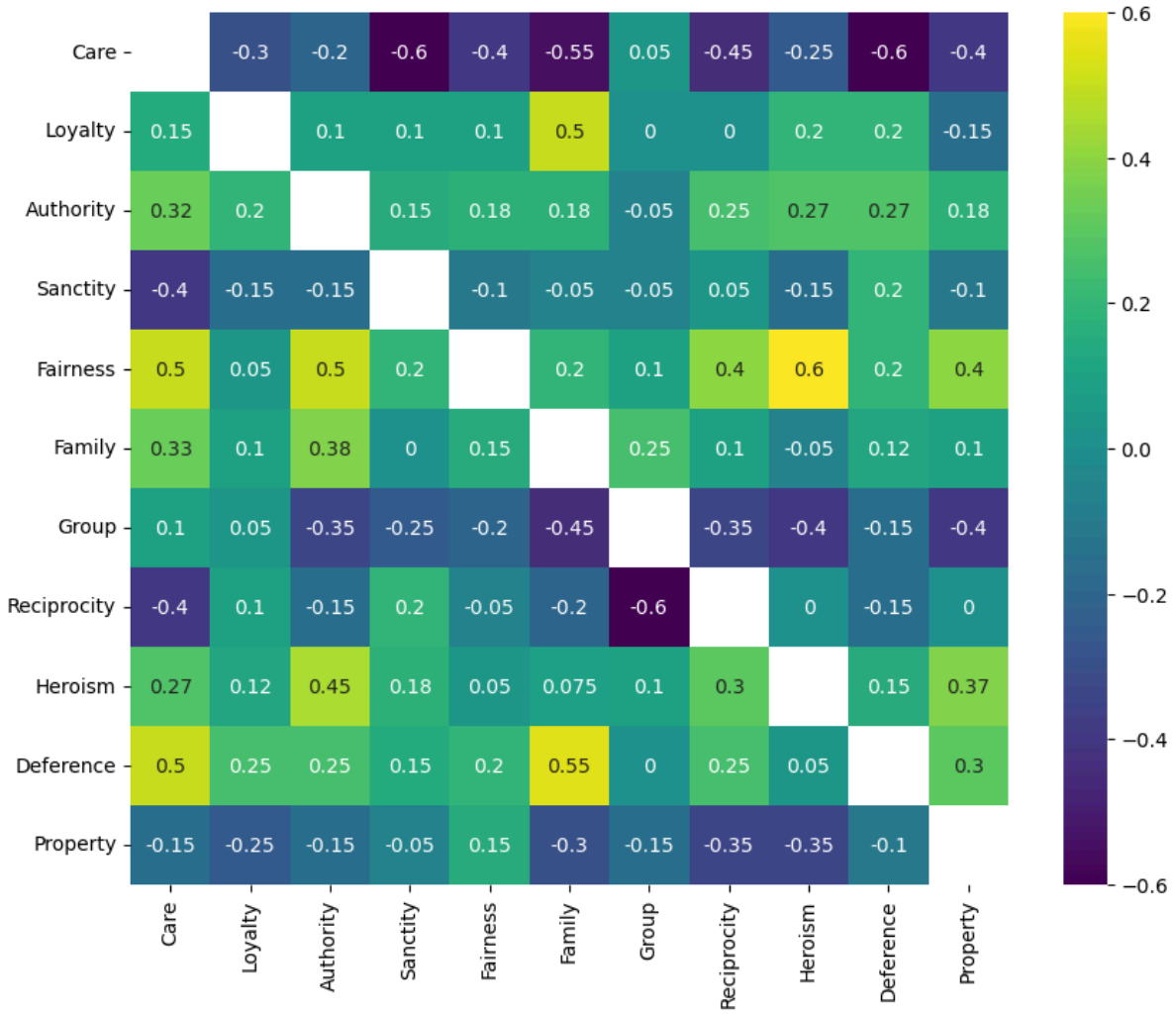
2. Comparing Neural Representations Across Moral Domains

For each moral domain, the accuracy of SVM was compared between those with high sensitivity and those with low sensitivity. The results indicated that sensitivity in care (t = -5.76, p = .0003), authority (t = 5.97, p = .0002), fairness (t = 5.28, p = .0005), group (t = -3.97, p = .003), heroism (t = 4.85, p = .0009), and deference (t = 4.56, p = .001) modulated the classification accuracy. Specifically, participants with high sensitivity in care (and group) demonstrated a more accurate classification of care (and group) vignettes from other moral domains. In contrast, the opposite pattern was observed for authority, fairness, heroism, and deference. Figure 11 provides the accuracy of the classifier for each pair of moral domains.

Addressing RQ6, SVM results indicate that the modulating effect of moral domain sensitivity is specific to the moral domain. Individuals with high sensitivity in care and group exhibit more distinct neural representations for these domains; hence, higher classification accuracy. In contrast, those with lower sensitivity in authority, fairness, heroism, and deference show more distinct neural representations of these moral domains.

Figure 11

Difference in Classification Accuracy Based on Moral Domain Sensitivity



Note. Each value indicates the difference in classification accuracy (i.e. accuracy for below median sensitivity - accuracy for above median sensitivity). Each row demonstrates the selected moral domain sensitivity. The columns indicate the moral domain of the vignette that was compared. For instance, the first row compared the participants with high sensitivity in care to those with low sensitivity. The second column of the first row provides the difference in classifying care vignettes with loyalty vignettes.

In order to examine the modulating effect of moral domain sensitivity on the distribution of neural representation of moral cognition, three variations in the measures of moral domain sensitivity were added to the linear combination of previous models. The linear combination of MA, MFT, H, and moral domain sensitivity scores with moral relevance measures was highest, accounting for 7.1% of the variance in the relative log Bayes factor in the free model (see Figure 12). The models that included moral domain sensitivity measure with the judgment section or the averaged moral domain sensitivity score accounted for 4.3% and 6.1% of the variance in the free model. To better understand the linear combination of the MAC + MFT + H + Moral Relevance model, the predicted second-moment matrix is provided in Figure 13. To answer RQ7, the results of PCM indicate that the representation model that incorporates moral domain sensitivity, particularly moral relevance measures for MAC, enhances the predictability of the pattern of neural activity across moral domains.

Figure 12

Standardized Log Bayes Factor for PCM

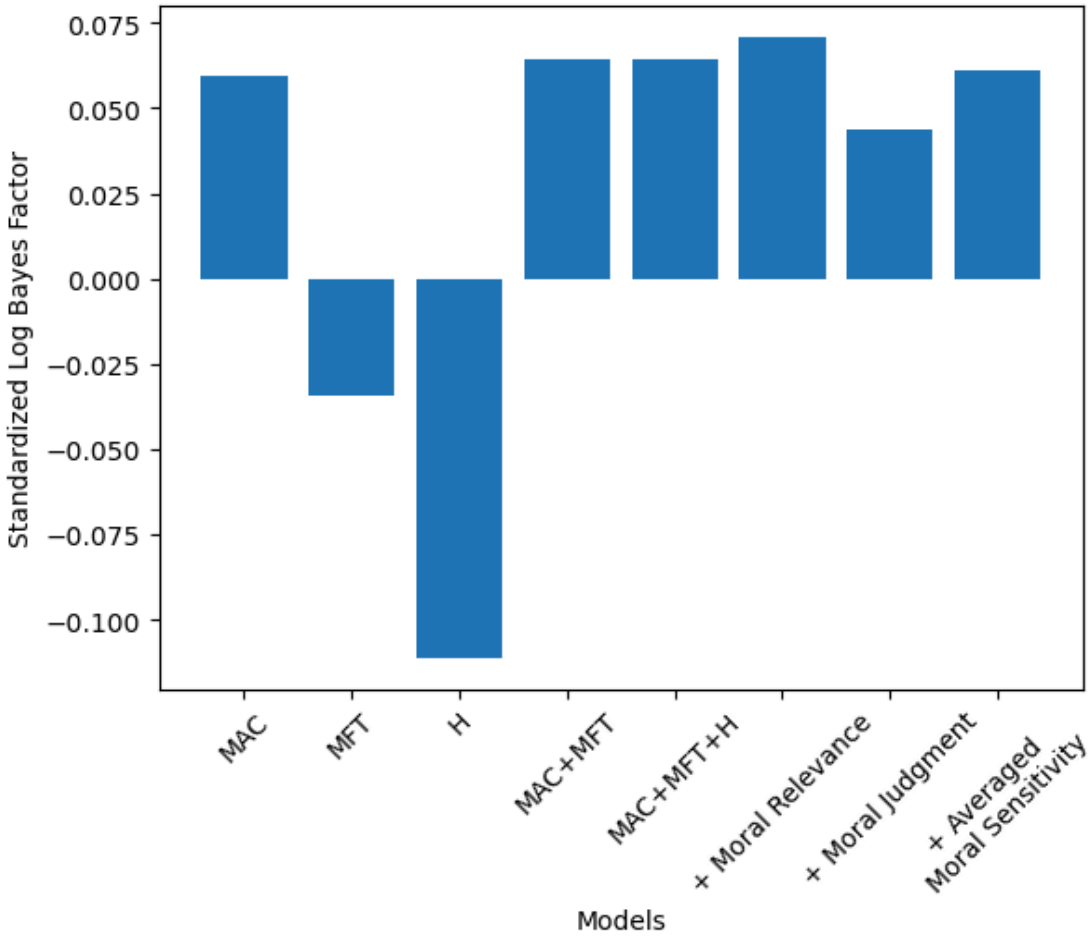
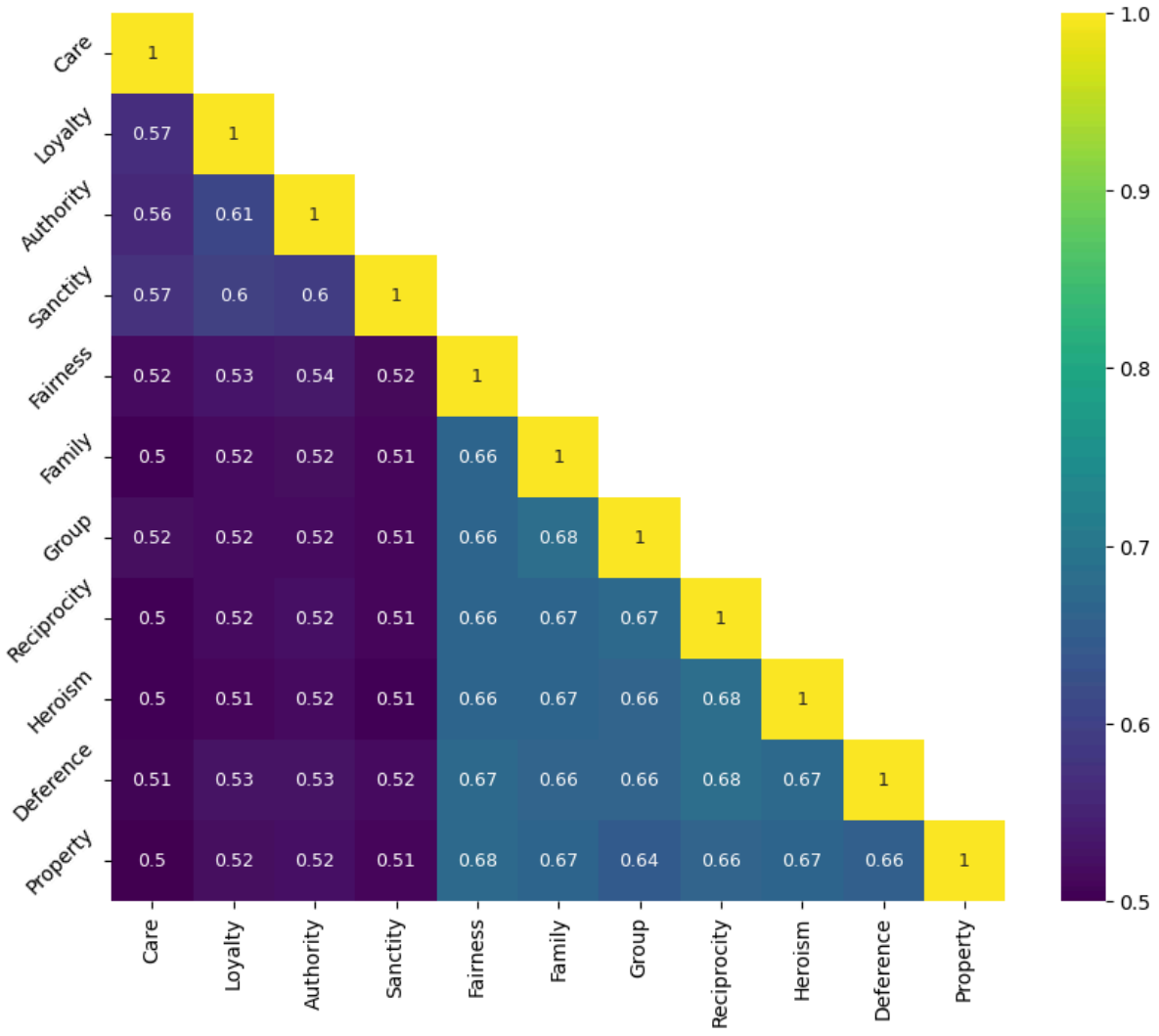


Figure 13

Predicted Similarity Matrix Using MAC + MFT + H + Moral Relevance Model



G. Robustness Checks

1. Identifying Overlapping Neural Representations of Moral Cognition

The robustness check, conducted through conjunction analysis, consistently aligns with the previous analysis, albeit yielding smaller clusters. Moral vignettes associated with MFT domains exhibited significant activation in brain regions, including the precuneus,

PCC, TPJ, superior temporal gyrus, and angular gyrus, even after controlling for potential confounding factors such as moral ratings, perceived believability, relatability, and severity (see Table 6 and Figure 14). Similarly, the conjunction analysis focusing on MAC domains substantiated the earlier findings (see Table 7 and Figure 15). A cross-domain comparison across all 11 moral domains identifies the precuneus as a significant cluster, with MNI coordinates of 0, -66, 42, a peak activation of 4.68, and a cluster size of 20 mm^3. This robustness check underscores the resilience of shared neural networks, particularly those associated with the theory of mind network, in moral cognition, even after controlling for other aspects of the moral vignettes.

Table 6

Results of Conjunction Analysis on MFT Domains

Cluster ID	MNI Coordinates			Peak Activation (Z)	Cluster Size (mm^3)
	X	Y	Z		
1	-0.12	-65.19	42.45	4.97	510
1a	-5.76	-59.54	33.98	3.85	
2	-58.40	-7.79	-12.13	4.75	446
3	53.46	-61.43	18.92	4.50	29
4	-0.12	-58.60	32.10	4.28	246
5	-49.94	-70.84	19.86	4.16	70
6	4.58	-56.72	21.75	3.89	18
7	-57.46	-67.07	17.04	3.80	24

Figure 14

Glass Brain Plotting of Conjunction Analysis on MFT Domains

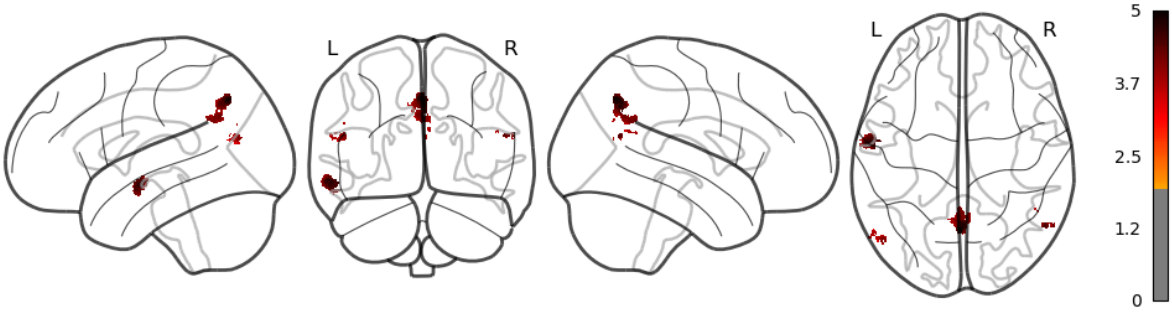


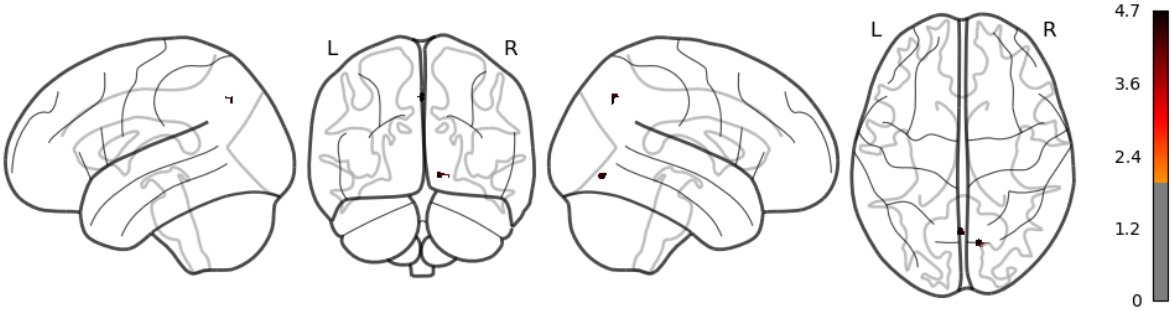
Table 7

Results of Conjunction Analysis on MAC Domains

Cluster ID	MNI Coordinates			Peak Activation (Z)	Cluster Size (mm^3)
	X	Y	Z		
1	0.82	-66.13	40.57	4.74	39
2	11.16	-72.72	-8.37	4.70	49

Figure 15

Glass Brain Plotting of Conjunction Analysis on MAC Domains



2. Comparing Neural Representations Across Moral Domains

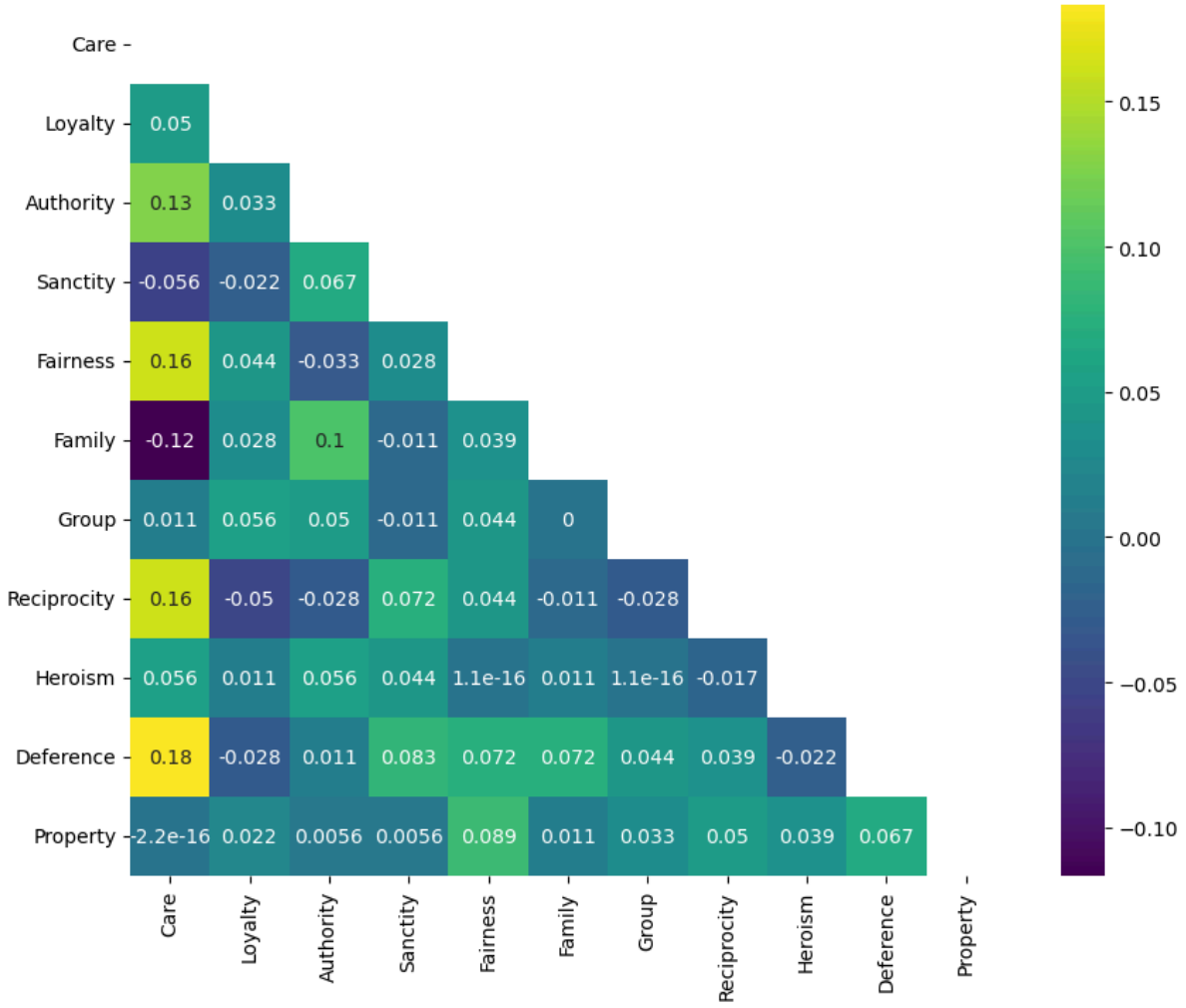
Overall, the classification accuracy results remained largely consistent with and without controlling for the evaluation of vignettes. The overall classification accuracy was 80.18% (*SD* = 19.88%), slightly lower than the previous result. The classification accuracy for moral domains within MFT also decreased to 83.94% (*SD* = 16.25%). Within MAC, the classification accuracy was 76.61% (*SD* = 22.29%), a 2.76% decrease from the previous result. The classification accuracy across theories (i.e., classifying one domain from MFT and the other from MAC) was also reduced to 81.73% (*SD* = 18.47%). Despite the decrease in accuracy after adjusting for individual differences in the evaluation of vignettes, the results still provide evidence that classifying moral domains from two different theories was above chance (see Figure 16 for the change in classification accuracies).

The robustness check results were mostly aligned with the previous findings on H4, H5, H6, and H7. Regarding H4, the classification accuracy of neural representation for care in MFT and family values in MAC was lower (75%) compared to the classification accuracy for these moral domains with the rest of the moral domains ($M = 81\%$), $t(17) = 2.90$, $p = .01$. Inconsistent with H5, the classification accuracy of neural representation for loyalty in MFT and group values in MAC was not lower (83%) than the classification accuracy for these moral domains with the rest of the moral domains ($M = 84\%$), $t(17) = 0.29$, $p = .83$. Partially consistent with the previous findings on H6, the classification accuracy of neural representations for authority in MFT and deference in MAC was marginally lower (67%) compared to the classification accuracy for these moral domains with the rest of the moral domains ($M = 73\%$), $t(17) = 2.09$, $p = .051$. Consistent with H7, the classification accuracy of neural representation for purity in MFT and other MAC domains ($M = 92\%$) was higher than

the classification accuracy for other MFT domains and MAC domains (M = 78%), $t(22)$ = 3.86, $p < .001$.

Figure 16

Change in Classification Accuracy of SVM for Pairs of Moral Domains



Note. A negative value indicates that the classification accuracy for the robustness check was higher than the previous results.

The PCM results underwent significant changes following the inclusion of vignette evaluations. Specifically, the H model, positing relationships between care-family,

loyalty-group, and authority-respect domains while assuming other domains to be equally dissimilar, accounted for 14.57% of the variance in the free model (refer to Figure 17). The corresponding predicted second-moment matrix is illustrated in Figure 18. In contrast to earlier findings, both the MAC model, the MFT model, and their linear combination now explained less variance than the null model, which assumes equal dissimilarity in neural representation across all 11 moral domains.

Figure 17

Standardized Log Bayes Factor for PCM

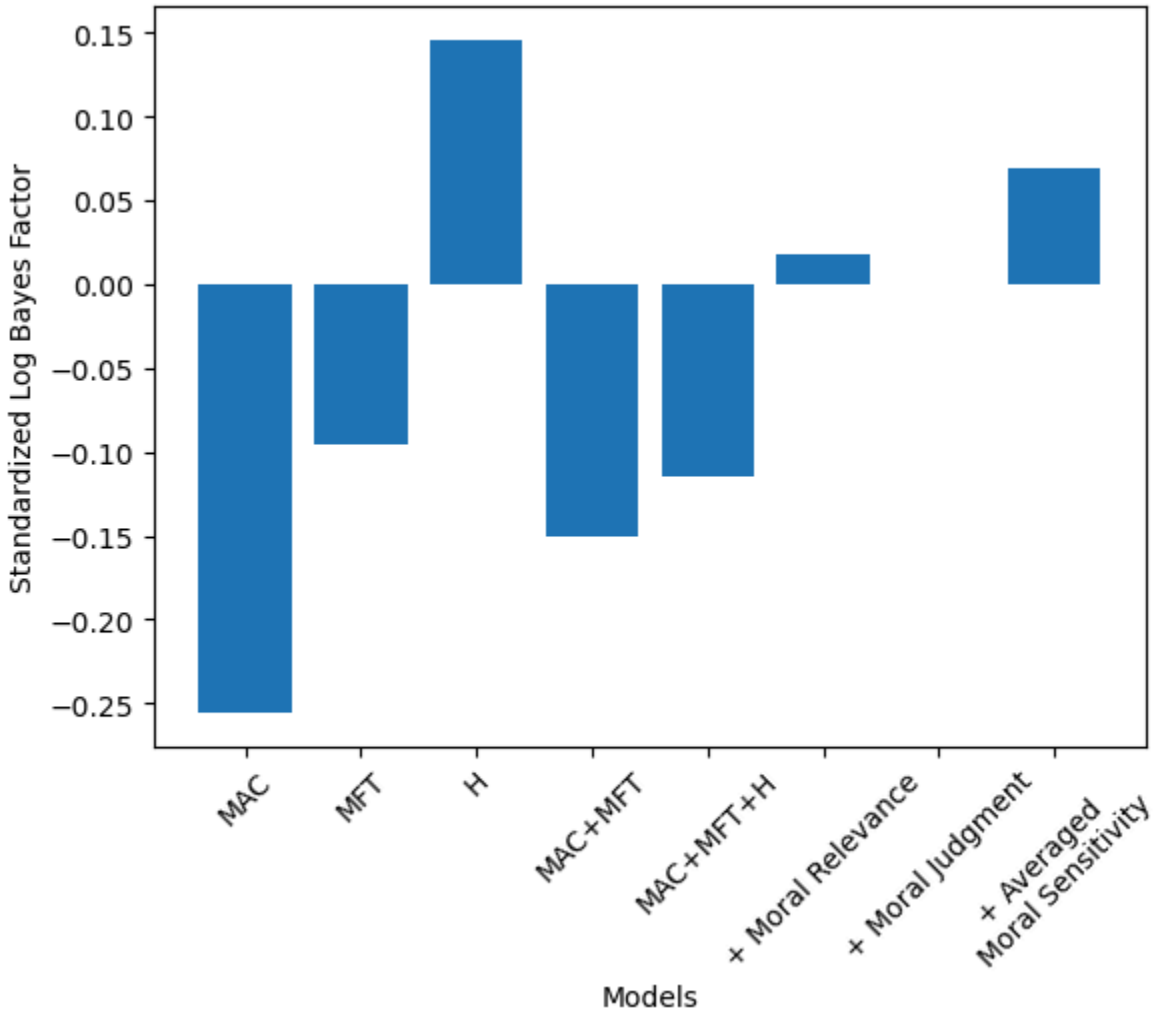
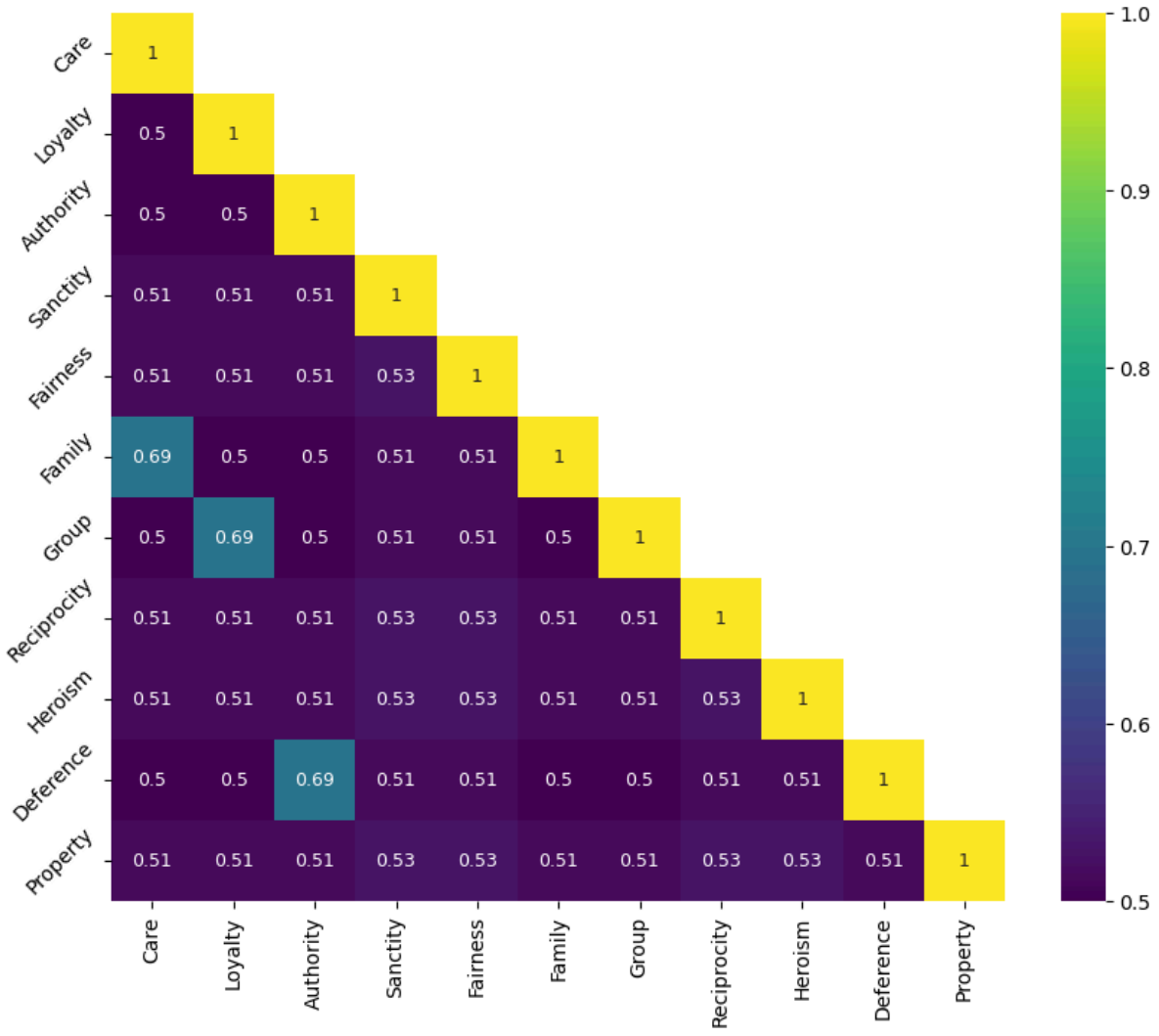


Figure 18

Predicted Similarity Matrix Using H Model



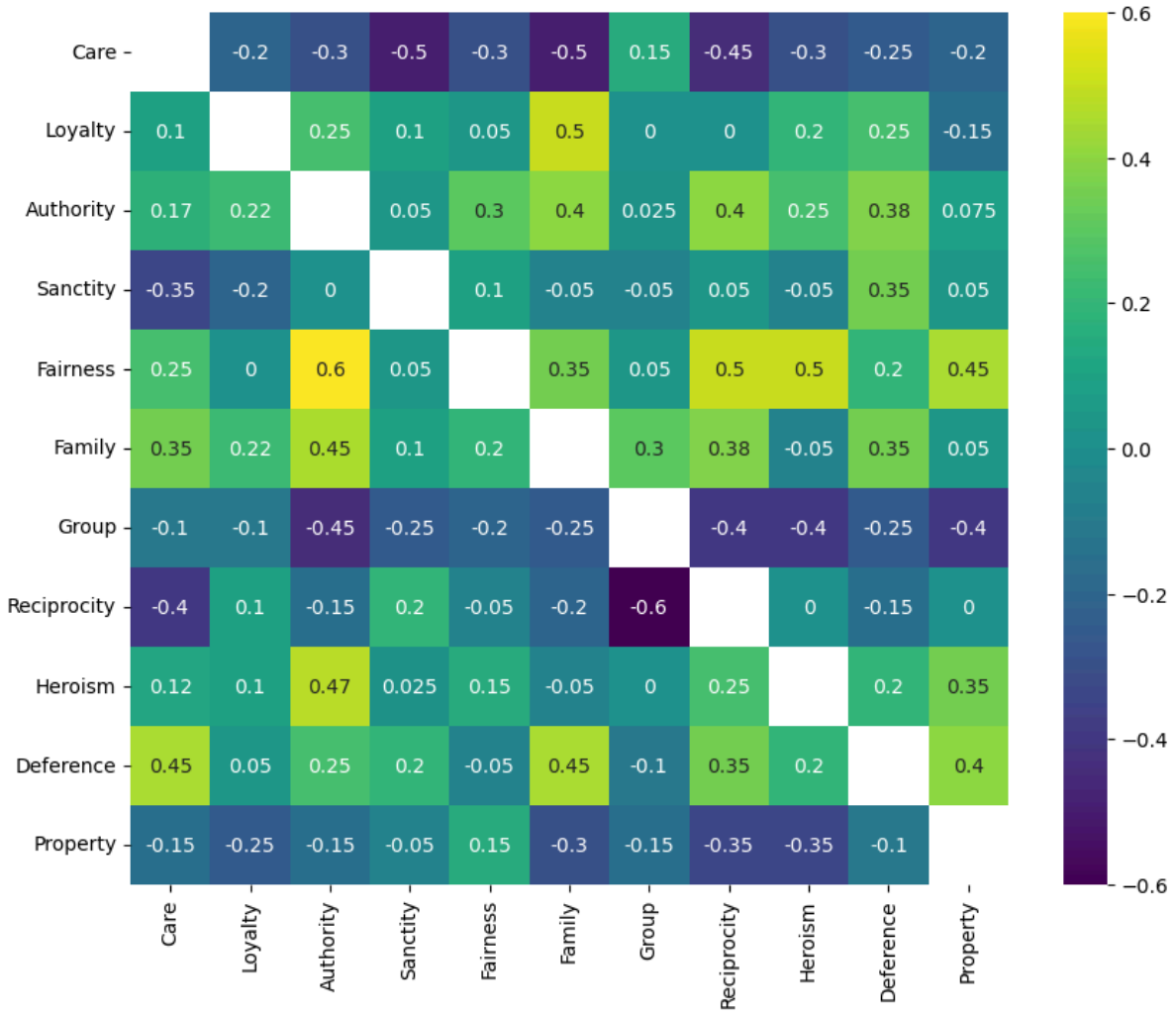
3. Examining Modulation of Moral Domain Sensitivity

In line with the prior conjunction analysis on moral domain sensitivity, no significant clusters were identified. Regarding the modulation of moral domain sensitivity on classification accuracy, results were consistent with the previous analyses for care (t = -4.74, p = .001), authority (t = 5.01, p = .0007), fairness (t = 4.03, p = .001), and group (t = -6.95, p

= .0006). However, the modulating effect of deference and heroism ceased to be significant when controlling for vignette evaluations (see Figure 19). Despite these differences, the overall classification accuracy results closely mirrored the previous findings, with a high Spearman correlation of .93 ($p < .0001$) between the classification accuracy matrices (Figures 11 and 19). Regarding PCM, the inclusion of moral domain sensitivity measures failed to explain the variance in the free model beyond the H model (see Figure 17), marking an inconsistency with the earlier findings.

Figure 19

Difference in Classification Accuracy Based on Moral Domain Sensitivity



Note. Each value indicates the difference in classification accuracy (i.e. accuracy for below median sensitivity - accuracy for above median sensitivity). Each row demonstrates the selected moral domain sensitivity. The columns indicate the moral domain of the vignette that was compared. For instance, the first row compared the participants with high sensitivity in care to those with low sensitivity. The second column of the first row provides the difference in classifying care vignettes with loyalty vignettes.

An exploratory analysis utilizing EFA on the predicted second-moment matrix of the free model (refer to Figure 20) discovered that a three-factor model accounted for 96.88% of the variance in neural representation. The factor loadings, detailed in Table 8, indicated that six out of seven MAC domains loaded on one factor, while loyalty in MFT, authority in MFT, and group in MAC loaded on the second factor. Additionally, care and sanctity loaded on the third factor. This observation signifies that the neural representation of the moral domains proposed by the two theories can be effectively distilled into three key factors.

Figure 20

Predicted Similarity Matrix Using the Free Model

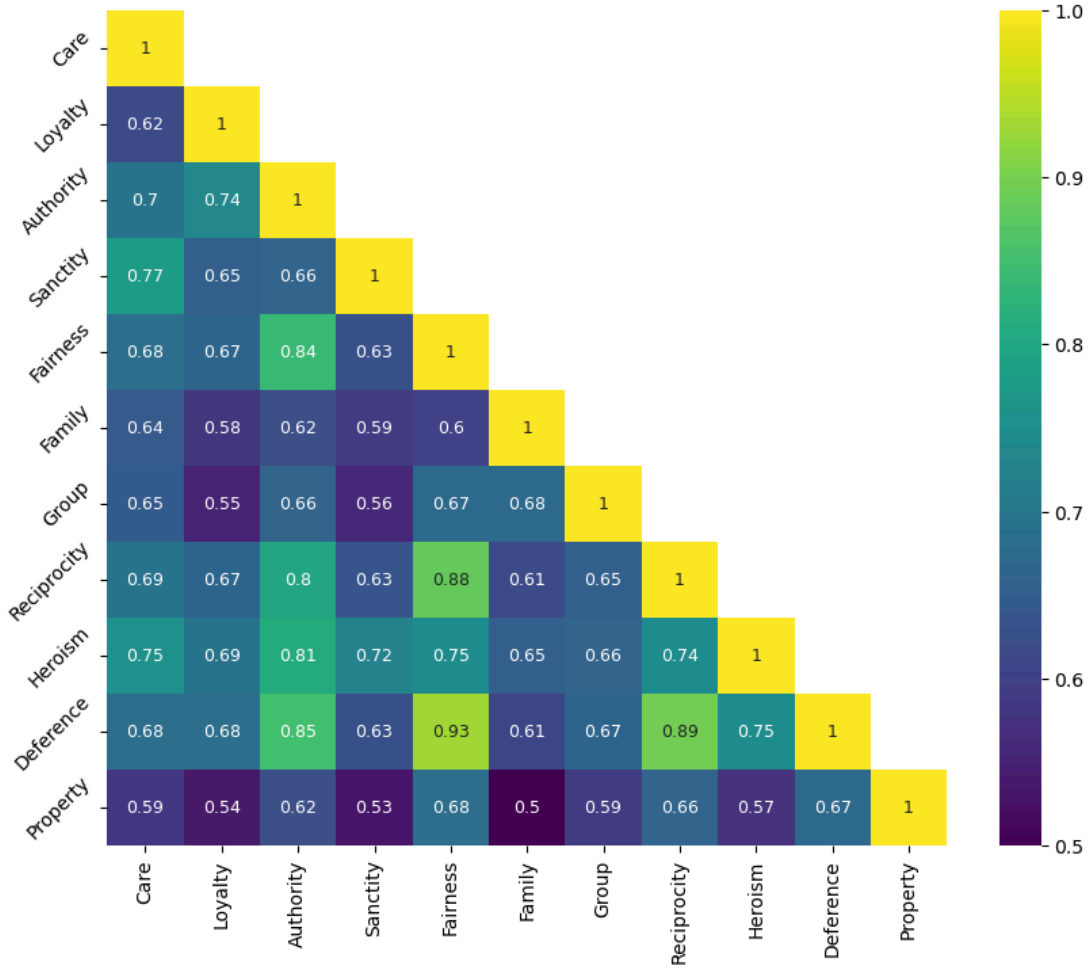


Table 8

Factor Loadings for EFA

	Factor 1	Factor 2	Factor 3
Care	-0.33	-0.09	**-0.94**
Loyalty	-0.28	**0.94**	0.10
Authority	0.21	**0.74**	0.58
Sanctity	-0.39	0.54	**-0.74**
Fairness	**0.84**	0.01	0.54
Family	**-0.89**	-0.41	-0.06
Group	-0.17	**-0.95**	0.21
Reciprocity	**0.79**	0.07	0.53
Heroism	**-0.78**	0.41	-0.36
Deference	**0.75**	0.07	0.66
Property	**0.94**	-0.26	0.21

V. Discussion

This dissertation addresses a critical gap in the moral cognition literature by undertaking a direct comparison between two prominent theoretical frameworks: MFT and MAC. MFT posits that morality is structured around a set of innate moral foundations, encompassing harm, fairness, loyalty, authority, and purity. Conversely, MAC introduces the concept of morality emerging from solutions to cooperation problems, with seven types of cooperation corresponding to seven types of morality. Despite their theoretical similarities and differences, there has yet to be a neuroscientific study directly comparing the two

theories. This dissertation, taking a neurological approach, offers new insights into how the proposed moral domains within and across the theories are related, providing theoretical and methodological insights to advance moral cognition research as well as implications for media studies.

A. Theoretical Implications

Aligned with prior research on MFT, the findings provide evidence that multiple neural networks play a role in MFT-associated moral cognition. Areas related to the theory of mind, sentence comprehension, and semantic processing are all involved in moral cognition, replicating previous literature with PCC, precuneus, and TPJ, as found by Greene and Haidt (2002) and Eres et al. (2018). However, only the precuneus emerges consistently across all moral domains for MAC. The specificity of this finding raises multiple questions. First, this observed distinction might be attributed to the methodology employed in the conjunction analysis, specifically focusing on brain areas associated with all seven moral domains in MAC. Alternatively, it could suggest inherent differences in the shared neural networks implicated in moral cognition between the two theories. Another plausible explanation is that the vignettes used in the study contribute to this divergence, shaping unique neural networks for each MAC domain or only a subset of them. Considering how all MAC domains are related to cooperative challenges, the moral vignettes provided more concrete situations that were easily believable and relatable for participants. This may have recruited a more specified neural network that is unique to each MAC domain (or only some of the MAC domains). However, for MFV, the vignettes were less relatable and believable, requiring more cognitive processes to understand and mentally taxing participants to imagine

themselves in the situations, as signified by larger activation in the theory of mind for processing MFV compared to MAC vignettes.

Survey measures and neural representations point to the likelihood that MAC domains will likely have one latent factor, argued by Curry (2016) to be the notion of cooperation, while MFT has two latent factors. Survey measures indicate that moral domains measured by MAC-Q, especially the moral relevance section, load on a single factor. However, moral domains measured with MFQ items load on two factors, with care and equality as one and loyalty, authority, and purity as the other. This is consistent with the previous literature on individualizing (i.e. care and fairness) and binding foundations (i.e. loyalty, authority, purity) of MFT (Napier et al., 2013). These three factors are also evident in the factor analysis of the predicted covariance of the neural representation using PCM, with few exceptions, such as group values not loading with other MAC domains but with MFT and sanctity loading with care instead of other binding foundations.

This dissertation also provides evidence for how the moral domains of the two theories are related. Firstly, consistent with Haidt and Joseph (2004) and Curry, Mullins et al. (2019), purity is an "odd corner" of morality. The classification accuracy for purity was significantly higher than other MFT domains. Second, the conceptual similarity in care-family, group-loyalty, and authority-respect is also somewhat supported at the cognitive level. When controlled for the believability, relatability, severity, and moral wrongness of moral cognition, the neural representation of these moral domains was more similar than other pairs of MFT and MAC domains. This suggests neurological evidence for the potential merging of these two theories.

B. Methodological Implications

The findings of this dissertation offer neurological evidence concerning how moral domain sensitivity for MAC should be measured. According to Curry et al. (2019), the moral relevance and moral judgment measures are not orthogonal but show significant discrepancies. Additionally, they found the moral relevance section of MAC-Q to be more reliable than the judgment section. This dissertation provides consistent neurological support for Curry et al.'s findings. The results of PCM demonstrated that moral domain sensitivity, as measured by the relevance section, outperformed the judgment section in predicting the neural representation of moral cognition.

This dissertation also highlights the need to revisit the vignettes created by Clifford et al. (2015). Although MFV was created to be used in neuroimaging studies, with social norms serving as a control condition, the social norms condition is confounded. Compared to vignettes describing moral transgressions, social norm vignettes were less relatable, less believable, and less likely to have severe consequences. While the results of GLM and MVPA were mostly stable with and without the addition of covariates to control for relatability, believability, and severity, analyses using the second-moment matrix, such as PCM, were highly volatile to these covariates. Therefore, the contrasts created by using social norms in MFV should not be considered the “best” control, and it should be noted that certain analyses are more likely to be influenced by these confounding factors.

C. Implications on Media Studies

By comparing the neural representation across moral domains, this dissertation provides evidence for moral pluralism and how individuals' moral domain sensitivity may

influence these moral domains. Considering that the conjunction analyses on the trend contrasts using moral domain sensitivity did not yield statistically significant clusters, and how the accuracy of pairwise domain classification using SVM varied across moral domain sensitivity, the findings suggest that the moral domains should not be treated equally. In other words, high sensitivity in some moral domains may result in a distinct neural representation, while low sensitivity in other moral domains may result in a distinct neural representation. Additionally, EFA on the free model of PCM also demonstrates that although family values, reciprocity, heroism, deference to property, and fairness loaded on one factor, family and heroism had negative loadings, suggesting an inverse relationship with the latent factor.

In the case of moral reframing, where messages that emphasize certain moral values can be rendered more compelling (Feinberg & Willer, 2019), its effect may vary depending not only on which moral domain the message receiver is sensitive to but also on which moral domain is being reframed in the message. Considering how those with high sensitivity to care demonstrated a more distinct neural representation of moral cognition compared to those with low sensitivity to care, using moral reframing on care may yield a more consistent effect on those with high sensitivity as opposed to those with low sensitivity. However, moral reframing on fairness values may solicit a more consistent effect on those with low sensitivity to group values than those with high sensitivity.

In media entertainment research, which has widely utilized MFT to better understand moral content in media (Tamborini, 2013), we may have only examined a few of the multiple moral domains. Considering how both survey, classification, and PCM results demonstrate a distinctive component for MAC, there may have been a large proportion of moral cognition and processes that current entertainment research needs to include.

Despite multiple moral research studies on social media, Neumann and Rhodes's review (2023) suggests that many publications were atheoretical. The findings of this dissertation provide strength to both MFT and MAC, as they are theoretically rooted in evolutionary and psychological approaches, supported by the neurological evidence presented. Furthermore, the theoretical arguments made by the two theories are broadly consistent when studying their neural representations, such as moral cognition for MFT loading on two factors and MAC loading on one factor, and each moral domain being distinctive enough, providing evidence for moral pluralism. Therefore, future research on social media can benefit not only from moral research but, more precisely, from theory-driven moral research.

D. Conclusion

This dissertation bridges a crucial gap in moral cognition literature by undertaking a comparative analysis of the MFT and MAC theoretical frameworks, shedding light on their respective neural representations. The theoretical implications underscore the distinctiveness of these frameworks in shaping moral cognition. Methodologically, the findings advocate for refined measures of moral domain sensitivity and prompt a reevaluation of vignettes, emphasizing the need for robust controls. This dissertation underscores the potential benefits of constructing vignettes with nuanced variations in severity, relatability, and believability for each moral domain. The implications for media studies underscore the importance of considering moral pluralism, suggesting that distinct neural responses exist based on individual sensitivities and that theoretical-driven approaches enhance our understanding of moral content in media and social platforms. Overall, this dissertation offers a nuanced

exploration of moral cognition, providing a foundation for future research endeavors in understanding the intricacies of morality in the human brain.

References

1. Amin, A. B., Bednarczyk, R. A., Ray, C. E., Melchiori, K. J., Graham, J., Huntsinger, J. R., & Omer, S. B. (2017). Association of moral values with vaccine hesitancy. *Nature Human Behaviour*, *1*(12), article 12. https://doi.org/10.1038/s41562-017-0256-5
2. Aquino, K., & Reed, I. I. (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
3. Aramovich, N. P., Lytle, B. L., & Skitka, L. J. (2012). Opposing torture: Moral conviction and resistance to majority influence. *Social Influence, 7*(1), 21–34. https://doi.org/10.1080/15534510.2011.640199
4. Atari, M., Graham, J., & Dehghani, M. (2020). Foundations of morality in Iran. *Evolution and Human Behavior*, *41*(5), 367–384. https://doi.org/10.1016/j.evolhumbehav.2020.07.014
5. Atari, M., & Haidt, J. (2023). Ownership is (likely to be) a moral foundation. *Behavioral and Brain Sciences*, *46*, article e326. https://doi.org/10.1017/S0140525X2300119X
6. Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S. T., & Dehghani, M. (2023). Morality beyond the WEIRD: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology*, *125*(5), 1157–1188. https://doi.org/10.1037/pspp0000470
7. Atari, M., Mostafazadeh Davani, A., & Dehghani, M. (2020). Body Maps of Moral Concerns. *Psychological Science*, *31*(2), 160–169. https://doi.org/10.1177/0956797619895284
8. Baez, S., Couto, B., Torralva, T., Sposato, L. A., Huepe, D., Montañes, P., Reyes, P., Matallana, D., Vigliecca, N. S., Slachevsky, A., Manes, F., & Ibanez, A. (2014). Comparing moral judgments of patients with frontotemporal dementia and frontal stroke. *JAMA Neurology*, *71*(9), 1172–1176. https://doi.org/10.1001/jamaneurol.2014.347
9. Barrett, H. C. (2012). A hierarchical model of the evolution of human brain specializations. *Proceedings of the National Academy of Sciences*, *109*, 10733–10740. https://doi.org/10.1073/pnas.1201898109
10. Behzadi, Y., Restom, K., Liau, 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
11. Binney, R. J., & Ramsey, R. (2020). Social Semantics: The role of conceptual knowledge and cognitive control in a neurobiological model of the social brain. *Neuroscience & Biobehavioral Reviews*, *112*, 28–38. https://doi.org/10.1016/j.neubiorev.2020.01.030

12. Brady, W. J., Wills, J. A., Burkart, D., Jost, J. T., & Van Bavel, J. J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, *148*(10), 1802.
13. 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.
14. Bruns, S., & Knop-Huelss, K. (2023). That's so immoral! Investigating the effects of moral violations reported in the form of (in) complete moral dyads in news articles on emotions and memory. *Human Communication Research*, *49*(1), 61-74.
15. Carvalho, F., Okuno, H. Y., Baroni, L., & Guedes, G. (2020, November). A brazilian portuguese moral foundations dictionary for fake news classification. In *2020 39th International Conference of the Chilean Computer Science Society (SCCC)* (pp. 1-5). IEEE.
16. Ciaramelli, E., Muccioli, M., Làdavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *2*(2), 84–92. https://doi.org/10.1093/scan/nsm001
17. Cingel, D. P., & Krcmar, M. (2020). Considering Moral Foundations Theory and the Model of Intuitive Morality and Exemplars in the context of child and adolescent development. *Annals of the International Communication Association*, *44*(2), 120–138. https://doi.org/10.1080/23808985.2020.1755337
18. Cingel, D. P., Krcmar, M., Marple, C., & Snyder, A. L. (2023). The development and validation of a measure of moral intuition salience for children and adolescents: The Moral Intuitions and Development Scale. *Journal of Communication*, *73*(2), 179–191. https://doi.org/10.1093/joc/jqac049
19. Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations vignettes: A standardized stimulus database of scenarios based on moral foundations theory. *Behavior Research Methods*, *47*(4), 1178–1198. https://doi.org/10.3758/s13428-014-0551-2
20. Curry, O. S. (2016). Morality as Cooperation: A Problem-Centred Approach. In T. K. Shackelford & R. D. Hansen (Eds.), *The Evolution of Morality* (pp. 27–51). Springer International Publishing. https://doi.org/10.1007/978-3-319-19671-8_2
21. Curry, O. S., Alfano, M., Brandt, M. J., & Pelican, C. (2022). Moral Molecules: Morality as a Combinatorial System. *Review of Philosophy and Psychology*, *13*(4), 1039–1058. https://doi.org/10.1007/s13164-021-00540-x
22. Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality*, *78*, 106–124. https://doi.org/10.1016/j.jrp.2018.10.008
23. Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate?: Testing the theory of Morality-as-Cooperation in 60 societies. *Current Anthropology*, *60*(1), 47–69. https://doi.org/10.1086/701478
24. Diedrichsen, J., Yokoi, A., & Arbuckle, S. A. (2018). Pattern component modeling: A flexible approach for understanding the representational structure of brain activity patterns. *NeuroImage*, *180*, 119-133.

https://doi.org/10.1016/j.neuroimage.2017.08.051
25. Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants?. *Trends in Cognitive Sciences*, 27(*7*), 597-600.
26. Eden, A., Tamborini, R., Grizzard, M., Lewis, R., Weber, R., & Prabhu, S. (2014). Repeated exposure to narrative entertainment and the salience of moral intuitions. *Journal of Communication*, *64*(3), 501–520. https://doi.org/10.1111/jcom.12098
27. 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
28. 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., Amlien, 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), article 7. https://doi.org/10.1038/s41596-020-0327-3
29. Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., & Bandettini, P. A. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, *215*, article 116828. https://doi.org/10.1016/j.neuroimage.2020.116828
30. Fu, W. W. (2013). National Audience Tastes in Hollywood Film Genres: Cultural Distance and Linguistic Affinity. *Communication Research*, *40*(6), 789–817. https://doi.org/10.1177/0093650212442085
31. Fulgoni, D., Carpenter, J., Ungar, L., & Preoţiuc-Pietro, D. (2016). An empirical exploration of moral foundations theory in partisan news sources. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 3730-3736. https://aclanthology.org/L16-1591
32. Gerbner, G., & Gross, L. (2000). Living With Television: The Violence Profile. In *The Fear of Crime*. Routledge.
33. Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Chapter Two - Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. In P. Devine & A. Plant (Eds.), *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Academic Press. https://doi.org/10.1016/B978-0-12-407236-7.00002-4
34. Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141
35. Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/10.1037/a0021847
36. Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, *6*(12), 517–523. https://doi.org/10.1016/S1364-6613(02)02011-9
37. 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
38. Guo, T., Wang, X., Wu, J., Schwieter, J. W., & Liu, H. (2024). Effects of contextualized emotional conflict control on domain-general conflict control: fMRI evidence of neural network reconfiguration. *Social Cognitive and Affective Neuroscience*, article nsae001.
39. Hahn, L., Tamborini, R., Prabhu, S., Grall, C., Novotny, E., & Klebig, B. (2022). Narrative media's emphasis on distinct moral intuitions alters early adolescents' judgments. *Journal of Media Psychology*, *34*(3), 165–176. https://doi.org/10.1027/1864-1105/a000307
40. Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*(4), 55–66. https://doi.org/10.1162/0011526042365555
41. Haidt, J., & Joseph, C. (2008). The moral mind: *How five sets of innate intuitions guide the development of many culture‐specific virtues, and perhaps even modules*. In P. Carruthers & S. Laurence (Eds.), *The Innate Mind, Volume 3* (1st ed., pp. 367–392). Oxford University PressNew York. https://doi.org/10.1093/acprof:oso/9780195332834.003.0019
42. Haidt, J., & Joseph, C. (2011). How Moral Foundations Theory succeeded in building on sand: A response to Suhler and Churchland. *Journal of Cognitive Neuroscience*, *23*(9), 2117–2122. https://doi.org/10.1162/jocn.2011.21638
43. Hayasaka, S., Peiffer, A. M., Hugenschmidt, C. E., & Laurienti, P. J. (2007). Power and sample size calculation for neuroimaging studies by non-central random field theory. *NeuroImage*, *37*(3), 721-730.
44. Hopp, F. R. (2021). Moral Idiosynchrony: Variability in Naturalistic Complexity Modulates Intersubject Representational Similarity in Moral Cognition [Ph.D., University of California, Santa Barbara]. In *ProQuest Dissertations and Theses*. https://www.proquest.com/docview/2629779760/abstract/3CDFF6E9068F4BFBPQ/1
45. Hopp, F. R., Amir, O., Fisher, J. T., Grafton, S., Sinnott-Armstrong, W., & Weber, R. (2023). Moral foundations elicit shared and dissociable cortical activation modulated by political ideology. *Nature Human Behaviour*, 1–17. https://doi.org/10.1038/s41562-023-01693-8
46. Hopp, F. R., Fisher, J. T., & Weber, R. (2020). Dynamic transactions between news frames and sociopolitical events: an integrative, Hidden Markov model approach. *Journal of Communication*, *70*(3), 335-355.
47. Iyer, R., Koleva, S., Graham, J., Ditto, P., & Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLOS ONE*, *7*(8), e42366. https://doi.org/10.1371/journal.pone.0042366
48. 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
49. Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143-156. https://doi.org/10.1016/S1361-8415(01)00036-6
50. Kennedy, B., Atari, M., Mostafazadeh Davani, A., Hoover, J., Omrani, A., Graham, J., & Dehghani, M. (2021). Moral concerns are differentially observable in language.

Cognition, *212*, 104696. https://doi.org/10.1016/j.cognition.2021.104696
51. Khoudary, A., Hanna, E., O'Neill, K., Iyengar, V., Clifford, S., Cabeza, R., De Brigard, F., & Sinnott-Armstrong, W. (2022). A functional neuroimaging investigation of Moral Foundations Theory. *Social Neuroscience*, *17*(6), 491–507. https://doi.org/10.1080/17470919.2022.2148737
52. Klebig, R. T., Matthias Hofer, Sujay Prabhu, Clare Grall, Eric Robert Novotny, Lindsay Hahn, Brian. (2019). The Impact of Terrorist Attack News on Moral Intuitions and Outgroup Prejudice. In *Media, Terrorism and Society*. Routledge.
53. Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the onion of brain representations. *Annual Review of Neuroscience*, *42*, 407-432.
54. 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
55. Lewis, R. J., Tamborini, R., & Weber, R. (2014). Testing a dual-process model of media enjoyment and appreciation. *Journal of Communication*, *64*(3), 397–416. https://doi.org/10.1111/jcom.12101
56. Martins, A. T., Faísca, L. M., Esteves, F., Muresan, A., & Reis, A. (2012). Atypical moral judgment following traumatic brain injury. *Judgment and Decision Making*, *7*(4), 478–487. https://doi.org/10.1017/S1930297500002813
57. Medin, D., Bennis, W., & Chandler, M. (2010). Culture and the Home-Field Disadvantage. *Perspectives on Psychological Science*, *5*(6), 708–713. https://doi.org/10.1177/1745691610388772
58. Moll, J., de Oliveira-Souza, R., & Eslinger, P. J. (2003). Morals and the human brain: A working model. *NeuroReport*, *14*(3), 299.
59. Mumford, J. A., Bissett, P. G., Jones, H. M., Shim, S., Rios, J. A. H., & Poldrack, R. A. (2023). The response time paradox in functional magnetic resonance imaging analyses. *Nature Human Behaviour*, 1-12.
60. Mumford, J. A., & Nichols, T. E. (2008). Power calculation for group fMRI studies accounting for arbitrary design and temporal autocorrelation. *Neuroimage*, *39*(1), 261-268.
61. Murray, D. R., & Schaller, M. (2016). Chapter Two - The Behavioral Immune System: Implications for Social Cognition, Social Interaction, and Social Influence. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 53, pp. 75–129). Academic Press. https://doi.org/10.1016/bs.aesp.2015.09.002
62. Napier, J. L., & Luguri, J. B. (2013). Moral mind-sets: Abstract thinking increases a preference for "individualizing" over "binding" moral foundations. *Social Psychological and Personality Science*, *4*(6), 754-759.
63. Neumann, D., & Rhodes, N. (2023). Morality in social media: A scoping review. *New Media & Society*. https://doi.org/10.1177/14614448231166056
64. Patriat, R., Reynolds, R. C., & Birn, R. M. (2017). An improved model of motion-related signal changes in fMRI. *Neuroimage*, *144*, 74-82. https://doi.org/10.1016/j.neuroimage.2016.08.051
65. Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage*, *84*, 320-341.

https://doi.org/10.1016/j.neuroimage.2013.08.048
66. Prabhu, S., Hahn, L., Tamborini, R., & Grizzard, M. (2020). Do morals featured in media content correspond with moral intuitions in media users?: A test of the MIME in two cultures. *Journal of Broadcasting & Electronic Media*, *64*(2), 255–276. https://doi.org/10.1080/08838151.2020.1757364
67. Prehn, K., & Heekeren, H. R. (2009). Moral judgment and the brain: A functional approach to the question of emotion and cognition in moral judgment integrating psychology, neuroscience and evolutionary biology. In J. Verplaetse, J. Schrijver, S. Vanneste, & J. Braeckman (Eds.), *The Moral Brain: Essays on the Evolutionary and Neuroscientific Aspects of Morality* (pp. 129–154). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6287-2_6
68. Rai, T. S. (2018). Relationship regulation theory. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 231–240). The Guilford Press.
69. Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*(1), 57–75. https://doi.org/10.1037/a0021867
70. Rueda, J. (2021). Socrates in the fMRI Scanner: The Neurofoundations of morality and the challenge to ethics. *Cambridge Quarterly of Healthcare Ethics*, *30*(4), 604–612. https://doi.org/10.1017/S0963180121000074
71. Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., ... & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage*, *64*, 240-256. https://doi.org/10.1016/j.neuroimage.2012.08.052
72. Singh, M., Kaur, R., Matsuo, A., Iyengar, S. R. S., & Sasahara, K. (2021). Morality-based assertion and homophily on social media: A cultural comparison between English and Japanese languages. *Frontiers in Psychology*, *12*, 768856.
73. Skurka, C., Winett, L. B., Jarman-Miller, H., & Niederdeppe, J. (2020). All things being equal: Distinguishing proportionality and equity in moral reasoning. *Social Psychological and Personality Science*, *11*(3), 374–387. https://doi.org/10.1177/1948550619862261
74. Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition, 131*(1), 159–171. https://doi.org/10.1016/j.cognition.2013.12.005
75. Suhler, C. L., & Churchland, P. (2011). Can innate, modular "Foundations" explain morality? Challenges for Haidt's Moral Foundations Theory. *Journal of Cognitive Neuroscience*, *23*(9), 2103–2116. https://doi.org/10.1162/jocn.2011.21637
76. Tamborini, R. (2011). Moral intuition and media entertainment. *Journal of Media Psychology*, *23*(1), 39–45. https://doi.org/10.1027/1864-1105/a000031
77. Tamborini, R. C. (2013). *Media and the Moral Mind*. Routledge.
78. Tamborini, R., Eden, A., Bowman, N. D., Grizzard, M., Weber, R., & Lewis, R. J. (2013). Predicting media appeal from instinctive moral values. *Mass Communication and Society*, *16*(3), 325–346. https://doi.org/10.1080/15205436.2012.703285
79. Tamborini, R., Hahn, L., Aley, M., Prabhu, S., Baldwin, J., Sethi, N., Novotny, E., Klebig, B., & Hofer, M. (2020). The impact of terrorist attack news on moral intuitions. *Communication Studies*, *71*(4), 511–527.

https://doi.org/10.1080/10510974.2020.1735467
80. Tamborini, R., Prabhu, S., Lewis, R. J., Grizzard, M., & Eden, A. (2018). The influence of media exposure on the accessibility of moral intuitions and associated affect. *Journal of Media Psychology*, *30*(2), 79–90. https://doi.org/10.1027/1864-1105/a000183
81. Tamborini, R., & Weber, R. (2020). Advancing the model of intuitive morality and exemplars. In K. Floyd & R. Weber (Eds.), *The Handbook of Communication Science and Biology* (pp. 456–469). Routledge.
82. Tamborini, R., Weber, R., Eden, A., Bowman, N. D., & Grizzard, M. (2010). Repeated exposure to daytime soap opera and shifts in moral judgment toward social convention. *Journal of Broadcasting & Electronic Media*, *54*(4), 621–640. https://doi.org/10.1080/08838151.2010.519806
83. Trepte, S. (2006). Social Identity Theory. In *Psychology of Entertainment*. Routledge.
84. Wasserman, E. A., Chakroff, A., Saxe, R., & Young, L. (2017). Illuminating the conceptual structure of the space of moral violations with searchlight representational similarity analysis. *NeuroImage*, *159*, 371–387. https://doi.org/10.1016/j.neuroimage.2017.07.043
85. Weber, R., Fisher, J. T., Hopp, F. R., & Lonergan, C. (2018). Taking messages into the magnet: Method–theory synergy in communication neuroscience. *Communication Monographs*, *85*(1), 81-102.
86. Weber, R., Mangus, J. M., Huskey, R., Hopp, F. R., Amir, O., Swanson, R., Gordon, A., Khooshabeh, P., Hahn, L., & Tamborini, R. (2018). Extracting latent moral information from text narratives: Relevance, challenges, and solutions. *Communication Methods and Measures*, *12*(2–3), 119–139. https://doi.org/10.1080/19312458.2018.1447656
87. Yi, D., & Tsang, J.-A. (2020). The relationship between individual differences in religion, religious primes, and the moral foundations. *Archive for the Psychology of Religion*, *42*(2), 161–193. https://doi.org/10.1177/0084672420909459
88. Yilmaz, O., Harma, M., & Doğruyol, B. (2021). Validation of Morality as Cooperation Questionnaire in Turkey, and its relation to prosociality, ideology, and resource scarcity. *European Journal of Psychological Assessment*, *37*(2), 149–160. https://doi.org/10.1027/1015-5759/a000627
89. 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
90. Youk, S., Malik, M., Chen, Y., Hopp, F. R., & Weber, R. (2023). Measures of argument strength: A computational, large-scale analysis of effective persuasion in real-world debates. *Communication Methods and Measures*, *0*(0), 1–23. https://doi.org/10.1080/19312458.2023.2230866
91. Young, L., & Dungan, J. (2012). Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience*, *7*(1), 1–10. https://doi.org/10.1080/17470919.2011.569146
92. Zakharin, M., & Bates, T. C. (2023). Moral Foundations Theory: Validation and replication of the MFQ-2. *Personality and Individual Differences*, *214*, 112339. https://doi.org/10.1016/j.paid.2023.112339