

UCLA

UCLA Electronic Theses and Dissertations

Title

The Role of Coherence in Reasoning about Social Issues

Permalink

<https://escholarship.org/uc/item/1w3676nx>

Author

Lee, Junho

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Role of Coherence in Reasoning about Social Issues

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Psychology

by

Jun Ho Lee

2022

© Copyright by

Jun Ho Lee

2022

ABSTRACT OF THE DISSERTATION

The Role of Coherence in Reasoning about Social Issues

by

Jun Ho Lee

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2022

Professor Keith J. Holyoak, Co-Chair

Professor Patricia W. Cheng, Co-Chair

Human reasoners assemble and connect a variety of information to arrive at a systematic understanding of the world. Coherence and parsimony are fundamental principles that guide this process by constraining the vast hypothesis search space. By taking note of the operations of these principles, we can better understand how real-world beliefs, judgments, and knowledge are formed and why some of them are highly resistant to change. In this dissertation, I present three research projects that involve laypeople's reasoning about several socially relevant topics. Each of the three projects may be viewed as a different manifestation of a core idea: human reasoners think and act in ways that maximize the perceived coherence between many psychological constructs, including their beliefs, motivations, experiences, internalized social norms, and a positive sense of self. Part 1 analyzes why it is difficult to persuade people holding entrenched beliefs, such as climate change skepticism. Then, a new intervention founded on the principles of coherence and parsimony is

tested for its effectiveness in promoting climate actions across the political spectrum. Inspired by the recent #MeToo movement, Part 2 examines the interplay between moral and non-moral judgments about professionals from different fields from the perspective of cognitive consistency theories. Analyses suggest that a decrease in people's evaluation of professionals in response to the professionals' moral transgression is driven by a decrease in their liking of the professionals. Part 3 investigates the influence of intergroup dynamics and identity on group-based moral judgments—more specifically, experience of collective guilt and pride, and assignment of responsibility to individuals following harmful or helpful interaction between groups. Two cross-societal experiments suggest that moral judgments made in realistic contexts align with social norms and individuals' identification with affiliated groups. Implications of these findings for social change and education are discussed.

The dissertation of Jun Ho Lee is approved.

Hongjing Lu

Dan Simon

Keith J. Holyoak, Committee Co-Chair

Patricia W. Cheng, Committee Co-Chair

University of California, Los Angeles

2022

TABLE OF CONTENTS

Introduction	1
Part 1: Promoting Climate Actions Using Cognitive Constraints	12
Study 1	13
Method.....	13
Results	22
Discussion	26
Part 2: Effect of Immoral Character on Non-Moral Judgments	30
Study 2	33
Method.....	35
Results	39
Discussion	44
Part 3: Collective Guilt, Pride, and Responsibility	46
Study 3A	50
Method.....	51
Results	55
Study 3B.....	60
Method.....	60
Results	65
General Discussion	72
Conclusion	76
References	81

LIST OF FIGURES

Figure 1.1. Diagram provided as feedback for the Tides exercise in Study 1	17
Figure 1.2. Graph for the first question in the climate-science exercise in Study 1	19
Figure 1.3. Probabilities of selecting political climate actions (1) in Study 1	23
Figure 1.4. Probabilities of selecting political climate actions (2) in Study 2	25
Figure 2.1. Ratings of competence, accomplishment, and contribution in Study 2	40
Figure 2.2. The mediated effect of immorality on ratings of targets' competence, accomplishment, and contribution in Study 2	42
Figure 3.1. Amount of money donated to O'Neal children's education in each condition in Study 3A	56
Figure 3.2. Participants' ratings that Brian should feel collective responsibility in Study 3A	57
Figure 3.3. Participants' experience of collective pride and feeling of contribution in Study 3B	66
Figure 3.4. Participants' experience of collective guilt, feeling of responsibility, and urge to apologize in Study 3B.....	67
Figure 3.5. Group-specific collectivism scores reported by participants from the U.S., South Korean, and Japan in Study 3B	70
Figure 3.6. Mediation analyses of the effect of culture on the dependent measurements in Study 3B	71

LIST OF TABLES

Table 1.1. Components of materials from experimental conditions in Study 1.	14
--	----

ACKNOWLEDGEMENTS

Before the end of my first month into the PhD program in the fall of 2016, I already realized how extremely fortunate I am to have some of the nicest mentors. The reason why I mention how nice Keith, Patricia, and Hongjing are before anything else is that I greatly respect them not only for how much they know (which goes without saying), but also for how wise and generous they are. Their patience and kindness had a way of making an international student from Korea feel comfortable in any settings. They always showed respect for students, graduate or undergraduate, were open to wild and silly questions, and encouraged us to explore new areas. I always think of them as my role models, not only because of their abundant knowledge but also because of their great characters. I will miss the Thanksgiving and Christmas dinners Keith and Hongjing invited me to.

I am very grateful that I had the opportunity to work with Dr. Keith Holyoak. He inspired and encouraged me to study moral psychology, which was the start of my academic journey in the U.S. He was willing to meet any time to have in-depth discussions and has shown me a very broad perspective on human cognition and artificial intelligence that I would not have been able to gain anywhere else. He is the most elegant and skilled writer I have ever met, and working on papers with him was a priceless learning experience.

I have grown a lot from conducting research with Dr. Patricia Cheng. Her uncompromising attention to details, enthusiasm for scholarship, and critical thinking are all invaluable assets for scientists that I hope to have too. Working with her and Emily on climate change skepticism opened the door to a whole new world for me, because it showed me how cognitive psychology theories can be used to solve complex but consequential real-world problems. Our project also had

a lasting impact on my career choice because I decided to delve deeper into research on social cognition and the intersection between cognitive science and social psychology.

Dr. Hongjing Lu was the first computational psychologist I worked with, and I have always been amazed by the variety of ways in which she can combine techniques and thoughts from statistics, engineering, and psychology. Her endless enthusiasm for finding new topics and domains of research was inspiring. I enjoyed working with her and Dr. P. Jeff Brantingham on the intergroup conflict project, which also led me to think more about how to utilize the findings in cognitive psychology to solve problems in social sciences. I would like to thank them for their considerate guidance.

I learned a great deal from other professors as well. The weekly CogFog hosted by Drs. Robert Bjork, Elizabeth Bjork, Barbara Knowlton, and Alan Castel had the most cheerful yet intense debates, and I will miss the intellectual stimulation from the meetings. Dr. Dan Simon's papers on coherent reasoning had a strong influence on my studies, so much so that I must have cited his work more than 30 times in this dissertation. Every one of his comments and suggestions was insightful and helpful in improving my dissertation.

I thank my dear friends I made in Los Angeles. Saskia Giebl, Akila Kadambi, Maureen Gray, Carolyn Murray, and Maggie Yeh in my cohort helped me so much in adjusting to this foreign environment from the first day. James Kubricht, Mac Xing, Mark Straccia, and Alex Siegel often took me out to have the least bit of socialization in my life, which I really appreciate. Emily Wong, Nick Ichien, J. Hunter Priniski, and Raihyung Lee are wonderful collaborators that I was very fortunate to have during my PhD years. Graduate school, especially when coupled with a global pandemic, seems to weigh heavily on students. But I know my lab-mates are some of the

most strong, kind, and intelligent people who will eventually get through the anxieties that we always shared, and do great things as PhDs in the future (get super rich, etc.)

Lastly, I thank my parents, my sister, and my partner Chaewon, who showed unconditional love and support. I would not have been able to finish this dissertation without you.

Vita

Education

- M.A. in Cognitive Psychology (2016), *Seoul National University*, Seoul, South Korea
- B.A. Cum Laude in Psychology (2013), Combined Minor: Brain-Mind-Behavior Minor, *Seoul National University*, Seoul, South Korea

Honors, Awards, & Fellowships

- Summer Mentored Research Fellowship, UCLA Graduate Division (07/2021)
- UCLA Cognitive Research Award, Rothman Gift Fund (07/2020)
- Graduate Summer Research Mentorship, UCLA Graduate Division (06/2017)
- Scholarship for Doctoral Study Abroad, Korea Foundation for Advanced Studies (09/2016 – 06/2022)
- National Research Fellowship for Humanities and Social Sciences, Korea Student Aid Foundation (03/2014 – 12/2014)
- Scholarship for Superior Academic Performance, Seoul National University (2011, 2012, & 2015)

Publications

- **Lee, J.**, Wong, E. F., & Cheng, P. W. (under review). Promoting climate actions: a cognitive-constraints approach.
- **Lee, J.**, & Holyoak, K. J. (2022). Familial guilt: A cross-society comparison of judgments of collective family responsibility. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni, (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*. Toronto, Canada: Cognitive Science Society.
- **Lee, J.**, Priniski, J. H., Valderrama, S., & Holyoak, K. J. (2021). Disgraced professionals: Revelation of immorality decreases evaluations of professional competence and accomplishment. In T. Fitch, C. Lamm, H. Leder, & K. Teßmar-Raible (Eds.), *Proceedings of the 43rd Annual Virtual Meeting of the Cognitive Science Society*. Cognitive Science Society.
- **Lee, J.**, & Holyoak, K. J. (2020). “But he’s my brother”: The impact of family obligation on moral judgments and decisions. *Memory & Cognition*, 48(1), 158-170.
- **Lee, J.**, & Holyoak, K. J. (2018). “But he’s my brother”: How family obligation impacts moral judgments. In C. Kalish, M. Rau, J. Zhu & T. T. Rogers (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- **Lee, J.**, Hyun, I., & Park, J. (2015). Categorization of academic papers by types of contributions: Comparison of *KJCBP* and *JEP: General*. *Korean Journal of Cognitive and Biological Psychology*, 27(4), 659-677.

Conference Presentations

- **Lee, J.**, Priniski, J. H., Valderrama, S., & Holyoak, K. J. (2021, July). Disgraced Professionals: Revelation of Immorality Decreases Evaluations of Professional Competence and Accomplishment. Poster presented at *the 43rd Annual Meeting of the Cognitive Science Society*. *Cognitive Science Society*.
- **Lee, J.**, Wong, E. F. & Cheng, P. W. (2020, July). Promoting pro-climate actions. Poster presented at *the 42nd Annual Virtual Meeting of the Cognitive Science Society*.
- **Lee, J.** & Holyoak, K. J. (2017). But he is my brother: The role of family obligation in moral reasoning. Poster presented at *the 58th Annual meeting of Psychonomic Society*, Vancouver, British Columbia, Canada.
- **Lee, J.** & Park, J. (2015). Unskilled but not necessarily unaware: Differentiating the effect of skill levels and task difficulty on overconfidence. Poster presented at *the 56th Annual meeting of Psychonomic Society*, Chicago, Illinois.
- **Lee, J.**, Hyun, I., & Park, J. (2015). Recent trends in KJCBP: Classification by types of academic contributions. Presented at *the 52nd annual meeting of Korean Society for Cognitive and Biological Psychology*, Jeju, South Korea.
- **Lee, J.** & Park, J. (2014). Unskilled, but not necessarily unaware. Poster presented at *the 55th Annual meeting of the Psychonomic Society*, Long Beach, California.

Other Presentations and Workshops

- Presentation at Stanford Psychology Causality in Cognition lab. “Asymmetry in Learning Common-cause vs. Common-effect Structures” (04/2022)
- Presentation at UCLA Psychology CogFog. “Judgment of Collective Responsibility” (05/2020)
- Workshop at UCLA Psychology MiniFog: R Tutorial (05/2020)
- Presented at UCLA Psychology Datablitz. “Mitigating Climate Change Skepticism: How to Talk (and Not Talk) to Skeptics, and Why” (01/2020)
- Presented at UCLA Psychology Datablitz. “Family Favoritism in Moral Dilemma” (10/2017)

Teaching

University of California, Los Angeles

- Laboratory in Cognitive Psychology (Psych 121), 10/2020 – 12/2020
- MATLAB Programming for Behavioral Sciences (Psych 20B), 04/2020 – 06/2020
- MATLAB Programming for Behavioral Sciences (Psych 20B), 01/2020 – 03/2020
- Laboratory in Cognitive Psychology (Psych 121), 04/2019 – 06/2019
- Introduction to Cognitive Science (Psych 85), 01/2019 – 03/2019
- Research Methods in Psychology (Psych 100B), 01/2018 – 03/2018

Introduction

In his book *Dialogue Concerning the Two Chief World Systems*, published in 1632, Galileo Galilei argued that the Copernican heliocentrism provides a better account of the universe and the planetary motions than the traditional Ptolemaic geocentrism. Even though astronomical observations were more elegantly and accurately explained by heliocentrism, Galileo was heavily rebuked by theologians and philosophers who abided by the biblical view of the universe. In the end, the Roman Catholic Inquisition persecuted Galileo of heresy and sentenced him to house arrest for the remainder of his life. The heliocentric view was regarded as blasphemous of humankind's relation to the order of the universe and a cardinal violation of moral values. It was not until 1992 that the Catholic Church officially vindicated him and admitted that his view was correct.

From a broader perspective, the two “world systems” that were at odds in this debate were not simply two ways of interpreting a set of data. Rather, the two models of the universe symbolized two different worldviews: one based on traditional Christian values and the other based on scientific principles (i.e., coherence and parsimony). Accepting a particular worldview means accepting an interlocking set of beliefs, attitudes, traditions, and moral values that constitute it. Because most of the people at Galileo's time were under the strong influence of the Christian worldview rather than a scientific one, they were reluctant to accept heliocentrism over geocentrism.

From today's perspective, it may be tempting to point out the seemingly irrational reasoning exhibited by people from the 17th century. However, this historical incident may have captured how human cognition reacts to any situation that creates a tension or inconsistency

between mental constructs (e.g., religious devotion vs. desire for a parsimonious explanation). Even in domains such as astronomy, where only dry facts and unsentimental logic seem to be of importance, our beliefs and actions are the product of a dynamic interplay between a variety of mental constructs.

Consider how our minds are inundated with a host of information, ideas, and desires related to countless topics in our daily lives. Sensory information of the shape of a cup, memory of yesterday's news on the war in Ukraine, motivation to appear intelligent and attractive at parties, and an indefinite number of other mental constructs implicitly and explicitly shape our thoughts and actions as we navigate in the world. While continuously exposed to such a massive amount of information, reasoners must still make decisions and tackle numerous practical problems to adapt and survive. How can we distinguish truths from falsehoods and rational arguments from irrational ones? How can we choose which pieces of evidence to connect to arrive at a useful understanding of the world? How can we think and act without being paralyzed by the sheer immensity of the available information?

One answer to these questions is *coherence*, a fundamental principle that has been hypothesized to guide human reasoning and information processing (Kant, 1781/1998; Thagard, 1989, 2002). In its most basic sense, coherence in cognition can be thought of as a rule that directs beliefs, judgments, and decisions that 'go together' (co-occur in most cases) to be activated simultaneously, but prohibits logically inconsistent or emotionally incompatible thoughts and actions from being elicited together (e.g., Simon et al., 2015; Spellman & Holyoak, 1992). This principle is a powerful tool used by reasoners to organize an immense amount of information systematically and efficiently, so that they can form generalizable explanations and narratives for future use.

However, despite sounding intuitive, coherence is a highly obscure concept that can carry substantially different meanings depending on which definition is used. Therefore, two different ways of conceptualizing coherence that have guided psychological theories will be briefly introduced for clarification. Then, how these approaches to coherence relate to the psychological experiments in this dissertation will be explained.

The first approach to coherence in reasoning focuses on the logical, explanatory relations between hypotheses and observed data. An older version of this view can be traced back to philosophical discussions about justification of truth (Davidson & LePore, 1986; Kant, 1781/1998). The framework of *explanatory coherence* (Thagard, 1989, 2002), for example, defines coherence in explanations as a formal relation between data, propositions (including interpretations of data), and hypotheses. Two propositions are considered as having an explanatory relation if one is a partial explanation of another, if they jointly explain another proposition or hypothesis, or if the two are analogous in their explanatory relations to a different pair of corresponding propositions. On the other hand, propositions are incoherent if they logically contradict each other. When hypotheses and propositions in a system cohere well with one another, that system is viewed as a logical, coherent explanation of a phenomenon. Importantly, Thagard and coherentists argue that this process is in fact used by reasoners to distinguish true propositions from false ones.

In the domain of causal learning, psychologists have similarly asserted that cognitive systems strive to maintain logically consistent and parsimonious sets of explanations for observed events in the world. Such explanations and causal models allow us to make accurate predictions and manipulate the environment to our advantage (Cheng & Lu, 2017; Powell et al., 2016). An explanation is considered more parsimonious when it can accurately account for more data using fewer assumptions. When people's prior explanations cannot explain new observations, reasoners

seek new explanations that exhibit greater explanatory coherence by accounting for a broader set of observations. This mechanism for rational updating of beliefs and explanations about the world is fundamental to our cognition, and holds regardless of the particular beliefs or attitudes people hold (e.g., their political orientation).

Although not specific to higher-order reasoning, the relationship between coherence, parsimony, and the problem of underdetermination in cognition can provide further insights relevant to the current discussion. The simple visual input from a 2-dimensional image of the Necker cube is easily understood by perceivers as a 3-dimensional object. However, even our perception of such a simplistic stimulus is not determined by the sensory data; there are infinitely many non-cubical 3-dimensional objects that can provide the same visual input as the Necker cube to our retina at a particular time and space (Pizlo, 2001). Therefore, to avoid paralysis, our cognitive system needs to solve this underdetermination problem by imposing constraints for the search in hypothesis space (Marr, 1982). In the case of perception, a powerful constraint is the default assumption that a distal object will have the simplest form out of all forms that are consistent with the input to our retina. This crucial step early in information processing is often overlooked in models of perception (e.g., Rumelhart et al., 1986) that assume that perceivers already have (and only have) the representations necessary for our typical perceptual experience (e.g., representations for equal-sized edges and areas of a cube).

This idea of necessary constraining of search space applies to cognition in general, and coherence and parsimony as principles in reasoning serve the same function: they constrain the vast search space for hypotheses that can explain data. In the social domain, we need to make sense of information from various sources by ‘connecting the dots’ ourselves and creating a good narrative. In an extreme case, conspiracy theories emerge as attractive explanations because they

provide a compelling story to explain numerous seemingly unrelated events by (overly) restricting the search space.

This first approach is founded on reasoners' aspiration for logical explanations and rational understanding of the world. Therefore, simple co-occurrence between beliefs or associations between psychological constructs that do not have causal or explanatory interpretations cannot be easily incorporated into the scope of these models. However, consider a scenario where a person has strong affection for a man named John but a dislike for wine. If the person learns that John likes wine, her affection for John may lead her to adjust the negative attitude toward wine. This example resonates with another line of psychological research on coherence that does not necessarily involve logic or causality.

The second approach to coherence stems from classical social psychology studies on cognitive dissonance and cognitive consistency theory. Researchers have underscored people's motivation to maintain internal balance between positive and negative attitudes (valences) associated with individuals, objects, and ideas (Festinger, 1957; Heider, 1946; Kunda, 1990). The core argument was that, when faced with inconsistency, people's beliefs and attitudes undergo dynamic shifts until a state of equilibrium is reached. In modern work, this process has been implemented as a constraint satisfaction network (Ditto et al., 2009; Holyoak & Simon, 1999; Kunda & Thagard, 1996; Simon et al., 2015; Thagard, 2008) that also incorporates aspects of 'hot cognition' such as motivation, liking, identity, and emotion as important factors. The positive or negative associations between these factors reflect logical as well as non-logical relations, including explanation, facilitation, and association (Thagard & Verbeurgt, 1998).

To illustrate, Kunda and Thagard (1996) proposed a parallel constraint-satisfaction model to explain how stereotypical beliefs can shift everyday perception of others' behaviors. In their

illustration, a person's implicit tendency to think of a black person as aggressive coheres (co-occurs) with the interpretation of his ambiguous action as aggressive. When given sufficient situational ambiguity to enable such coherence shifts, people even arrive at different factual judgments that better align with conclusions supported by prior beliefs and motivations (Alicke, 2000; Lee & Holyoak, 2019; Simon et al., 2015; Ditto et al., 2009; Thagard, 2008).

Social psychology studies on cognitive consistency theory have often viewed this coherence shift in a negative light. It was argued that our tendency to achieve and maintain a coherent view of the world results in a trade-off between accuracy and efficiency in information processing, eventually biasing our reasoning. Moreover, it is unclear how beliefs and attitudes arise in the first place and how positive and negative associations between them solve the underdetermination problem. In contrast, another recent extension of the coherence-based reasoning framework to the moral domain (Holyoak & Powell, 2016) takes advantage of the constraint-satisfaction model. Deontic rules such as rights, duties, norms, and obligations are essential to describing our everyday moral judgments and decisions, but modern theories of moral psychology were not able to supply representations of those concepts. Under this view, rights and duties are effectively encoded as pragmatic reasoning schemas (Cheng & Holyoak, 1985; Holyoak & Cheng, 1995) that are part of constraint-satisfaction network models and constrain behaviors, judgments, and beliefs.

To avoid confusion, I will refer to the interpretation of coherence suggested in the first approach as *rational coherence* and the interpretation from the second approach as *cognitive consistency* for the rest of the introduction section. The two interpretations have some similarities in their roles, such as allowing people to organize their beliefs and experiences in systematic ways to make predictions and solve problems. Thagard and Verbeurgt (1998) attempted to unify these

views under ‘coherence’ as an umbrella term. They further laid out a list of domains and tasks in which coherence is used by reasoners, such as assessment of truth, ethical justification, legal justification, practical reasoning, perception, and impression formation. Kruglanski (1990) proposed a similar view, positing that knowledge formation and revision in any social-cognitive domain involves hypothesis formation and validation, regardless of whether logical (e.g., causal attribution) or non-logical (e.g., attitudes, stereotypes) relations are used. However, the two views still have meaningful distinctions, including the kinds of inputs they operate on (hypotheses and data vs. a variety of constructs relevant to both ‘cold’ and ‘hot’ cognition), and the problems they typically aim to solve (acquiring accurate explanation and prediction vs. satisfactory perception, judgment, and decision making). Therefore, in building psychological models, it is still useful to be mindful of the two broad approaches to coherence, and how they map onto different types of psychological tasks.

It should be noted that the goal of this dissertation is not to pit one interpretation of coherence against another, or to challenge the basic idea that coherence plays a critical role in human reasoning. My goal is to highlight the practical significance of these differing interpretations of coherence and show how these theories can be independently or jointly applied to investigations of laypeople’s social cognition.

Rational coherence and cognitive consistency can jointly help us understand how people form particular beliefs and judgments in the real world and why they often refuse to change them. In Part 1 of this dissertation, I analyzed the entrenched beliefs of climate change skeptics in the U.S. and tested the effect of an intervention that is designed to mitigate the skepticism. This intervention was informed by both rational coherence and cognitive consistency. Based on cognitive consistency theories (Ditto et al., 2009; Kunda & Thagard, 1996), we can infer that

people's polarized opinions about political issues are likely embedded in an intricate network of related beliefs, emotions, past experiences (e.g., of negative interaction with a member of an opposing party), and salient identities (e.g., as a proud Christian and a patriot). Therefore, presenting a few pieces of isolated statistical evidence from scientific papers to conservatives will not be sufficient for bringing change to this stable network.

How then can we change people's entrenched beliefs in a reliable way? Because the theories of rational coherence provide insights into how our causal knowledge is revised, the design of intervention for promoting climate actions could benefit from those theories. More specifically, the intervention in Study 1 simulated the rational process of causal knowledge revision in the following steps: (1) experience of deviation between data and predictions from prior causal model (violation of causal invariance, Cheng & Lu, 2017), (2) comparison between prior explanation and a more coherent explanation for the observation, and (3) acceptance of a new explanation. This process of knowledge revision was repeated at the level of individual tasks; however, the intervention also attempted to construct a coherent, moral, pro-science narrative across tasks that participants can readily adopt regardless of their political identities.

In Part 2, I investigated how moral motivations and social perceptions with conflicting valences interact from the perspective of cognitive consistency theories. Participants were asked to make specific social judgments (ratings of competence, contribution to society, etc.) about fictitious professionals from different fields before and after learning about a severe moral transgression they had committed (multiple sexual assaults or drug-related murder). The goal was to test if participants adjust their non-moral evaluations of professionals in response to the professionals' moral transgressions, even though the transgressions were technically orthogonal to their proficiency or competence. Participants were hypothesized to experience imbalance between

positive perception of professionals' competence and their strong immoral character. Cognitive consistency theories (Haidt, 2012; Heider, 1946; Kunda & Thagard, 1996) predict that this inconsistency will be resolved to some extent through dynamic shifts in judgments. However, a similar previous study (Stellar & Willer, 2018) proposed a rational pathway (decrease in assessment of social intelligence) through which evaluation of moral character influences non-moral judgments. In the present study (Study 2), I used mediation analyses to show that the effect is better accounted for by additionally using a non-logical variable: participants' psychological involvement with the target (or likability of target).

Lastly, in Part 3, I examined how group-based moral judgments made in first-person perspective are influenced by realistic factors such as intergroup dynamics and group identity. Moreover, I discussed the possibility of tight vs. loose social norms (Gelfand et al., 2011) playing an additional role on top of those individual factors when collective responsibility is assigned from a third-person perspective. Results showed that collective moral judgments made in realistic contexts indeed correlate with individuals' identification with affiliated groups. These findings can be better interpreted from the perspective of cognitive consistency, because a set of diverse constructs and deontic rules (including internalized social norms, sense of obligation to make reparations, and group identities) were seen as constraining participants' moral judgments.

In summary, all studies throughout this dissertation attempt to view our reasoning processes in various social contexts through the lens of coherence. Each of the three projects may be viewed as a different manifestation of a core idea: human reasoners think and act in ways that maximize the perceived coherence (both rational coherence and cognitive consistency) within an ensemble of beliefs, desires, motivations, and even a positive image of self (Alicke & Sedikides, 2009). Taken together, these studies aim to offer a richer understanding of the processes through

which people form, maintain, and change their beliefs and decisions about contentious issues in the world.

Reasoning about social issues in today's world requires gathering and organizing an overwhelming amount of complex and noisy data from numerous sources. Without coherence and parsimony as the guiding principles that help us create our own hypotheses and narratives to explain the events in the world, we may be lost in the flood of those information, unable to make sense of the data. However, those same principles may lead us to form entrenched beliefs and self-perpetuating narratives (e.g., conspiracy theories) that are resistant to persuasion using any sound evidence.

Understanding the conditions that make it difficult to change people's minds can inform us about effective ways to change their opinions and decisions about consequential issues in the world. Especially in our post-truth era, objective standards for discerning truth are fading, and the increasing political and economic polarization (e.g., Bail et al., 2018) makes social consensus more difficult to achieve. In this light, both rational coherence and cognitive consistency can be extremely useful. Cognitive consistency can provide analyses of conditions that perpetuate undesirable beliefs and actions (e.g., especially influential and persistent beliefs or attitudes in a network), as it did in all three parts of this dissertation. Theories of rational coherence can provide guidance on how to design educational materials that can systematically change people's entrenched beliefs.

Part 1: Promoting Climate Actions Using Cognitive Constraints

Despite Earth’s climate emergency (Masson-Delmotte et al., 2018), few countries are committed to meeting the goals set in the Paris Climate Agreement (Kennedy, 2020). The majority of people in the U.S. and U.K. now believe that ACC is under way (Crawley et al., 2020; Leiserowitz et al., 2020), and about 70% of Americans—the middle four segments of “global warming’s six Americas” (Leiserowitz et al., 2018), with views toward ACC ranging from “concerned” and “cautious” to “disengaged” and “doubtful”—are open to shifting their beliefs and actions toward pro-climate governmental policies. Yet, less than half of the American voters responded that they currently prioritize governmental climate policies when casting their votes in elections. Without a strong and unified public demand, governments are unlikely to implement the concerted changes necessary in energy infrastructure and laws, and inaction will result in environmental changes that are beyond humanity’s control.

Unfortunately, the political divide on the judgments and decisions related to ACC is salient. When Leiserowitz et al. (2020) asked respondents to choose important issues that will determine their vote, global warming was ranked as the 15th out of 28 issues. However, liberal Democrats (4th) and conservative Republicans (28th) showed a drastic difference in their priorities, and the intermediate groups between the two extremes did not show a strong interest either (moderate/conservative Democrats: 16th, liberal/moderate Republicans: 23rd). Similarly, Leiserowitz et al. (2018) showed that there was about 30% gap between liberals and conservatives in their willingness to take political and social actions to mitigate climate change (e.g., donating money to and volunteering for organizations working on global warming, urging politicians and government officials to focus on climate change).

ACC is a domain in which systematic misinformation disseminated by the U.S. fossil fuel industry has seriously impeded social progress (McKibben, 2019). As a result of decades of lobbying and advertising campaigns, climate change skeptics have formed conspiratorial narratives that provide their own coherent explanations for evidence of climate change and climate scientists' account of it. On the one hand, the narratives adopted by climate skeptics function by providing logical accounts of social and scientific phenomena (e.g., the corrupt climate scientists want to profit from environmental policies; the current increase in global temperature is part of the natural fluctuation found throughout Earth's history; the peer-review process in scientific journals is biased). On the other hand, the endorsement of climate change skepticism is closely intertwined with the protection and embracement of conservative identities for many Republican voters (e.g., Democrats want to drive the conservative workers in the fossil fuel industry out of jobs; environmental regulations encroach American citizens' freedom; Hochschild, 2018).

Just as the shift between the two conceptions of planetary motion involved a view of the self and of moral values within a system of beliefs that went beyond interpretations of astronomical data, persuading conservatives to take climate action cannot be simplified to conveying more scientific knowledge about climate change. Previous surveys (Drummond & Fischhoff, 2017; Kahan et al., 2012) have shown that the polarization of beliefs toward ACC is more salient in Americans who scored higher in science knowledge and numeracy (proficiency in understanding quantitative information) compared to Americans who scored lower in those measurements. In this light, an effective persuasion should provide conservative Americans with a pro-scientific, pro-environmental narrative to replace their previous narratives and worldviews, while minimizing the perceived threat to their self-identity (Steele, 1988). This study aimed to design and test a set of exercises that can jointly convince liberal as well as conservative participants.

Study 1

In this study, a variety of activities and educational information were woven into a coherent, moral, pro-science, and pro-environmental narrative. On a broad level, three modules were developed and named *moral identity*, *coherence-of-science*, and *climate change* modules, respectively. The *moral identity* module was expected to invoke participants' sense of moral self and elicit positive image of the intervention and materials as a whole, so that the information in other modules can be easily accepted (Kunda & Thagard, 1996). The *coherence-of-science* module was designed to highlight how scientific explanations provide a coherent and parsimonious account of various phenomena in our world (Lombrozo, 2016; Thagard, 1989). The *climate change* module introduced scientific data from NASA as evidence that ACC is taking place. The conjunction of the first two modules was expected to enable an inoffensive, rational, and moral persuasion that successfully encourages even conservative participants to take pro-climate actions. On the other hand, the *climate change* module was included to test if a similar effect can be obtained by directly conveying scientific knowledge and data about ACC.

In essence, this study aimed to demonstrate the importance of recognizing the targets of persuasion as moral and rational agents. That is, their attitudes and decisions were expected to change through coherence-based reasoning regardless of their political orientations and beliefs on ACC.

Method

Participants and Design

A total of 459 participants (253 females, 3 non-binary gender, $M_{age} = 37.4$, $SD_{age} = 12.4$) were recruited through Amazon Mechanical Turk. There were 19 additional participants who were removed due to responses that were inattentive (e.g., wrote nonsensical responses such as “good”

when asked to explain their answers) as judged by two independent scorers. The scorers resolved the few differences.

Previous research (Levay et al., 2016) showed that there is a substantially smaller proportion of Republicans on MTurk (14.6%) than in a more representative dataset (American National Election Studies 2012, 29.4%). To have maximal access to conservative participants on MTurk, participants were recruited from only the ten U.S. states that scored lowest in the belief that climate change is human-caused (Howe et al., 2015). In this dataset, 33.8% self-identified as moderate or extreme conservative, 42.4% as moderate or extreme liberal, 20.3% as independent, and 3.4% as libertarian.

Of the recruited participants, 94 independents and 17 libertarians were excluded, leaving 348 (198 females, 3 non-binary gender, $M_{age} = 38.3$, $SD_{age} = 12.6$), in order to enter political identity as a continuous predictor (from *extreme conservative* to *extreme liberal*) into regression analyses. Independents were expected to exhibit high variability in their political opinions, and it was unclear where the libertarians would stand on the conventional political spectrum with respect to ACC (Kahan et al., 2012).

Table 1.1. Component materials in each of six experimental conditions in Study 1.

Condition	<i>Moral Identity</i>	<i>Coherence of Science</i>	<i>Climate Change</i>	Assessment survey
a) Full coherence ($n = 84$)	√	√	√	√
b) All but climate ($n = 31$)	√	√		√
c) All but moral ($n = 34$)		√	√	√
d) Moral only ($n = 42$)	√			√
e) Climate only ($n = 34$)			√	√
f) Control ($n = 86$)				√

The experiment included 6 conditions that varied in the presence and absence of the three intervention modules (see Table 1). Although conducting a three-way full factorial analysis will require 8 conditions in total ($2 \times 2 \times 2$), the current study focused on testing the synergistic effect of *moral identity* and *coherence-of-science* modules, with and without *climate change* module. Presenting all three modules (condition *a*) was expected to maximize the coherence of the pro-scientific, pro-environmental, moral narrative proposed in the experiment. Therefore, participants' willingness to take climate actions was expected to be the highest in condition *a*. More generally, the intervention effect was predicted to be large when *moral identity* and *coherence-of-science* modules were included in the intervention (conditions *a* and *b*). On the other hand, presenting those modules one by one (conditions *d* and *e*) was not expected to yield such a great effect compared to the control condition.

About twice as many participants were recruited for the full-coherence and control conditions (conditions *a* and *f*) as for the rest because they were part of both the 1) *moral-identity* \times *coherence-of-science* design and 2) *climate-science* \times complementary-coherence (two other modules combined) design. Depending on the condition, the medians of the time spent on the entire survey ranged from 5.1 to 37.9 min. All participants were paid at the rate of around \$8 per hour.

Materials and Procedure

The opening instruction to participants stated, "Your opinions and thoughts are important to us, because a major goal of our study is to understand how we as people think about issues." The wording connotes that the researchers are in the same inclusive group as they the participants. To avoid proselytizing, consistent with the respect for participants' independent thinking, the constructed modules did not appear related, leaving to participants the task of "connecting the

dots”. To avoid direct clashes with participants’ preconceptions, the intervention materials made no mention of “climate change” or other politicized terms throughout.

In the *moral identity* module, participants read an article featuring a clinical psychologist’s interview with Iowan farmers who became suicidal upon facing severe financial difficulty from the Midwestern floods (Rossman & Atkin, 2019). Participants were asked to put themselves in the shoes of the farmers who called the therapist for help. They then chose what kind of help they would like to receive from the therapist among the provided options (e.g., “Help me apply for crop insurance”, “Give me information on how to contact FEMA for emergency financial assistance”). They were also asked to suggest other important means of addressing the issue and to explain their answers briefly. By doing so, this module was intended to set an empathic moral stance that persists throughout the experiment (Ochs & Capps, 2001).

Next, participants were shown a set of seven general-science prediction exercises (*coherence-of-science* module) and one climate-science prediction exercise (*climate science* module) in a fixed order (the climate-science exercise appeared fifth). All eight prediction exercises shared a common structure: A problem regarding what the outcome would be in a real-world phenomenon (e.g., water poured out from drinking glasses) or a familiar scientific invention (GPS, antibiotics), with a question asking participants to predict the outcome and to give their best explanation for their prediction. The subsequent feedback showed the actual outcome (accompanied by a video when appropriate), a science-based explanation for the correct answer, and relevant references (e.g., URLs for NASA, NOAA, and other relevant educational websites). By inducing participants to generate their own explanations (Chi et al., 1994) and juxtapose them with more coherence scientific explanations of the same phenomenon, the science prediction tasks aimed to highlight the parsimony in scientific explanations (Thagard, 1989) and the value of scientific worldview.

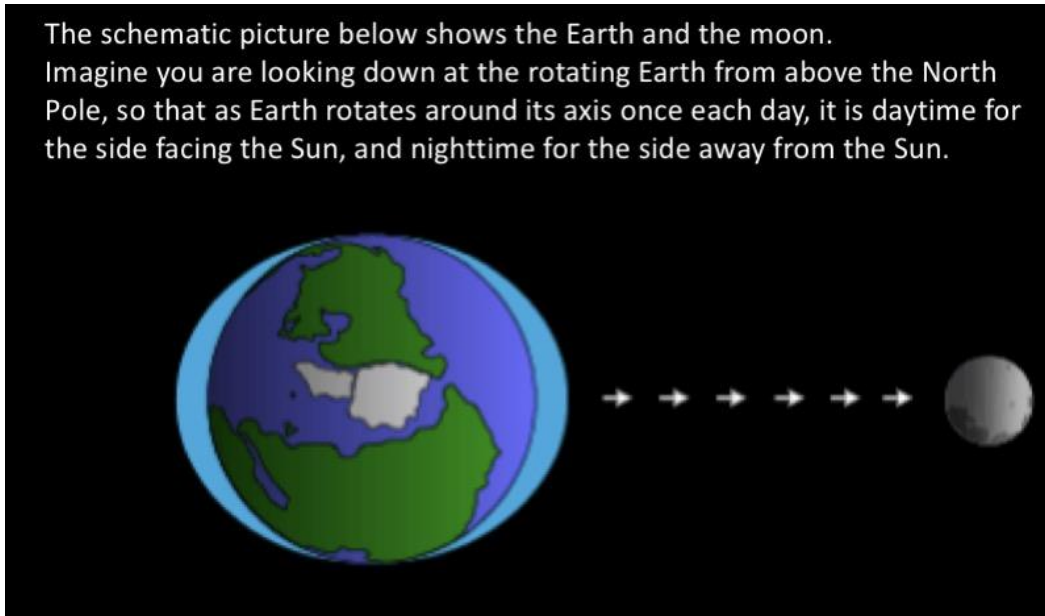


Figure 1.1. The diagram provided as part of the feedback for the Tides exercise.

For brevity, one of the general-science prediction exercises and the climate science prediction exercise will be presented here.

Tides Exercise.

Participants were asked, “Gravitational force from the moon causes tides in coastal areas. How many high tides and low tides do you think a coastal area experiences a day? (Tides are regular rises and falls of sea levels.)” A video showed high and low tides at the Bay of Fundy. Participants typed in the number of high and low tides in two blanks and wrote the explanations for their answers.

The feedback stated that there are two high tides and two low tides a day, noting, “Whether you be a fisherman or a surfer, the time for each of the 4 tides on a day is what you will find on any tide table.” It included a diagram (see Figure 1), along with an explanation in terms of Newton’s law of universal gravitation:

As Earth rotates a full circle once a day (counter-clockwise in the diagram), the moon's gravity pulls on different parts of our planet. By Newton's law, for any two bodies, the closer they are to each other, the stronger the gravitational pull between them. Accordingly, **the part of the ocean nearest to the moon is pulled the most**, resulting in a **bulge** (see diagram).

By the same law, **the part of the ocean farthest away from the moon is pulled the least**, resulting in **another bulge** (think about it as being left behind the most). ...

So, as your part of the coast on Earth **rotates into** these **two bulges** of water, **once each day**, you experience a **high tide**. And as your part rotates into **each area in between** the two bulges, because water is drained away from you (toward the bulges), you experience a **low tide**.

The question continued,

By the way, what if the **same natural force** that causes the tides **were to stop**? What path do you think our moon would take? Earth's massive pull on the moon is what keeps it from flying off, causes it **to curve** its path, keeping it close, as we see crossing our night sky every night.

The following link was provided as a source of scientific explanation: <https://scijinks.gov/tides/>. Most participants predicted incorrect numbers of tides. When asked to explain their answers, many of the participants either wrote that they had guessed, simply restated that there will be one high tide and one low tide, or reasoned that the tide will be high when the moon is close to the coastal area at night and low during the rest of the day. The feedback in the next page then showed participants how theories developed by scientists can advance our coherent understanding of the world.

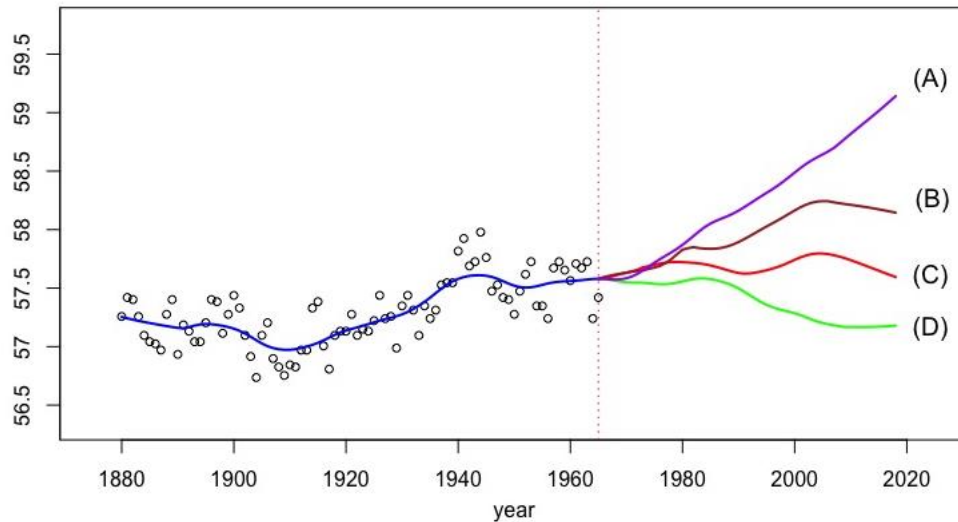


Figure 1.2. Graph showing the first question in the climate-science exercise.

Climate-Science Exercise.

Participants were first asked to project a trend of an unknown quantity “ x ” shown on a global-temperature graph from the year 1880 to 1965 (see Figure 2). The label for the ordinate (i.e., global temperature) was notably omitted. The projection therefore could be based on only the visual past trend. Participants chose from four response options, ranging from various continuations of the previous trend with little or no change to a salient upward rise (see the curves labeled from “A” to “D” in Figure 2). Up to this point, the question did not involve climate.

The goal of the “unknown x ” projection exercise was to let participants commit to what they considered a natural project, one that deviates from the actual observed trend. On the next page, ‘ x ’ was revealed to be the observed average global surface temperature from 1880 to 2018 (reported by NASA, <https://climate.nasa.gov/vital-signs/global-temperature/>). A new graph similar to Figure 2 displayed the correct label for the ordinate and the temperature data from 1965 to 2018, which aligned with curve (A). Only few participants (11%) chose the correct answer.

The deviation of participants' own judgment from the actual observed trend was expected to elicit the need to revise their causal understanding of the phenomenon. That is, the causal invariance constraint (Cheng & Lu, 2017), which people assume by default, was hypothesized to compel participants to suppose a new cause of that deviation in order to produce a coherent explanation of the data.

Accordingly, the follow-up question asked participants to explain what they thought caused the substantial recent deviation from the trend preceding 1965. The feedback page provided additional data graphs in support of ACC: the graphs showed the rapid rise in global surface temperature co-occurring with the sharp rise in human emission of greenhouse gases, but not with levels of other causes such as volcanic activity or solar activity (Masson-Delmotte et al., 2013).

The next page sketched the explanation of a broad range of climate phenomena by climate science, illustrating the parsimony of the explanation. It highlighted various outcomes predicted by the rise in temperature according to climate scientists' models:

Long before our current experience of extreme weather patterns and global temperature rise, climate scientists have predicted these outcomes. For example, in 1981, ... atmospheric physicist James Hansen, son of a tenant farmer in Iowa, predicted that the 21st century would see more extreme droughts and floods as climate zones shift, the erosion of ice sheets resulting in worldwide rise in sea level, and the opening of the fabled Northwest Passage.... these impacts predicted decades ago ... have since eerily become facts.

Note that this feedback information provided a potential explanation for the extreme flooding inflicting farmers in the Midwest in the *moral identity* module.

The page then showed graphs of simulated global temperature on land and ocean based on models using only natural forcings and models using both natural and anthropogenic forcings, with the observed temperature clearly aligning with the latter (IPCC, 2007).

Lastly, the materials included a figure showing 97%-consensus among climate scientists (Lewandowsky et al., 2013). The consensus information has been shown to be effective for raising acceptance of ACC. However, if distrust of science is a reason for resisting climate actions, conveying consensus among climate scientists would not raise willingness to take climate actions. Including that information in the climate exercise allowed a sufficiency test: if the climate-only condition fails to raise willingness to take climate actions compared to the control condition, then that information, when embedded in other materials, does not suffice to raise that willingness.

Assessment Survey and Dependent Measurements.

All conditions ended with a survey consisting of several sets of questions that measured the demographics and dependent variables. My dissertation will focus on three key dependent variables: 1) agreement with general science-attitude statements, which were selected from a handbook on the test of science-related attitudes (Fraser, 1981; e.g., “I am curious about the world in which we live”, “Science helps to make life better”; rated on a scale ranging from 1: strongly disagree to 5: strongly agree), 2) willingness to take two political climate actions (“Participate in a climate-action demonstration”, “Vote for legislations that promote policies that help create a more sustainable planet”; binary Yes/No responses were collected, scores were integers between 0, 1, and 2), 3) three science-fact knowledge questions, selected from a survey by the Pew Research Center (Funk et al., 2019):

Q1. What is the main cause of seasons on Earth? (Correct answer: The tilt of the Earth’s axis in relation to the plane of its orbit around the Sun),

Q2. When large areas of forest are removed so land can be converted for other uses, such as farming, which of the following occurs? (Correct answer: Increased erosion),

Q3. An antacid relieves an overly acidic stomach because the main components of antacids are? (Correct answer: Bases).

Demographic questions asked for age, gender, education level, political identity.

Results

As a reminder, this study aimed to test whether more coherent narratives have greater influence on the willingness to take climate actions. Because the willingness to take climate actions was measured as a discrete variable that takes values of zero, one, or two, ordinal logistic regression models were used to estimate the probabilities of observing each of those values.

First, the willingness to take political climate actions was predicted by 1) the presence/absence of the *moral identity* module (1: present, -1: absent), 2) the presence/absence of the combination of two other components (*coherence-of-science* module and *climate change* module; 1: present, -1: absent), and 3) the two-way interaction between the first two factors, while controlling for political orientation, education level, and the level of science-fact knowledge.

As Figure 1.3 shows, a significant increase in the willingness scores was observed when the complementary coherence questions and the *moral identity* module were presented together (condition *a*). Neither the farmers question alone, $OR = 0.53, p = .098$, nor the science exercises alone, $OR = 0.53, p = .12$, significantly changed the odds of selecting an additional political climate action relative to the control condition. With the farmers question present, however, the science exercises nearly sextupled the odds of selecting an additional action, $OR = 5.98, p < .001$. Conversely, with the science exercises present, the farmers question more than tripled the odds of selecting an additional action, $OR = 3.13, p = .006$. The interaction between the farmers and science

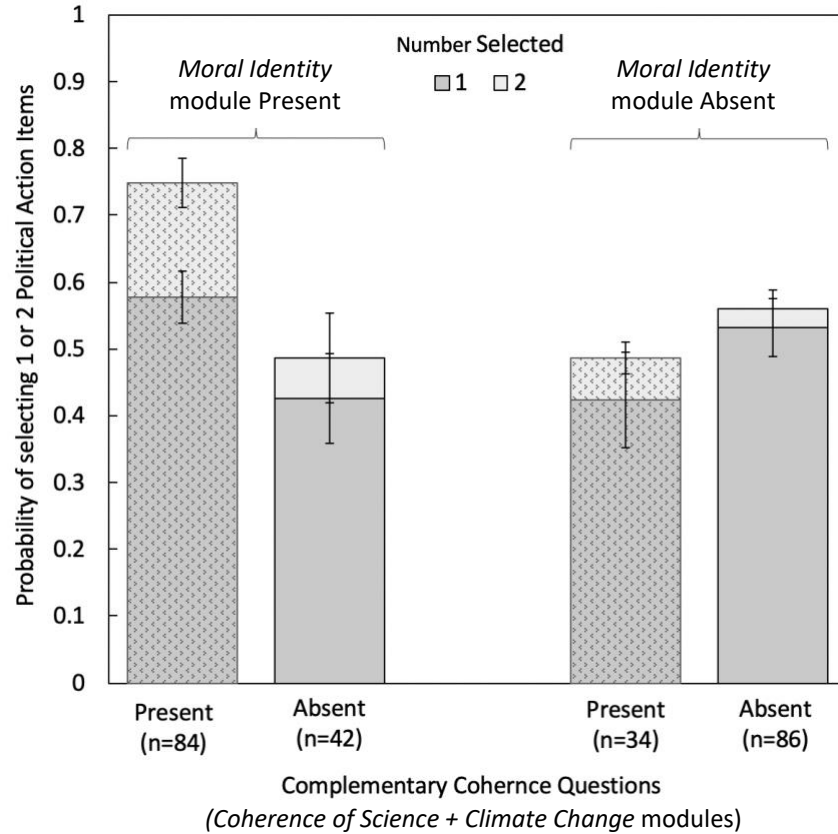


Figure 1.3. Probabilities of selecting zero, one, or two political climate actions, estimated by the presence/absence of the *moral identity* module and the presence/absence of the other two modules (four bars from left to right correspond to conditions a, d, c, and f from Table 1, respectively). ‘Complementary Coherence Questions’ here denote the combination of the *coherence of science* module and the *climate change* module. The unmarked height above each bar indicates the estimated probability of selecting zero action.

exercises was highly significant, $OR = 1.79, p = .002$. The mutual enhancement of the effects by the farmers and science questions strongly supports the cognitive-constraints hypothesis.

Liberals and conservatives did not significantly differ in the interaction between farmers and complementary coherence, $OR = 0.41, p = .44$. This result was obtained by a three-way

factorial ordinal logistic regression on the frequencies of selecting zero, one, or two political climate actions with the farmers question, the complementary-coherence questions, and binary political identity (conservative vs. liberal) as fixed factors, controlling for science-fact knowledge and educational level by incorporating them as covariates. Compared to the *moral identity* only condition, adding the complementary-coherence questions raised the probability of selecting political climate actions for both conservatives and liberals ($ps < .03$; one-tailed).

Next, the willingness score was predicted by 1) the presence/absence of the *climate change* module (1: present, -1: absent), 2) the presence/absence of the combination of two other components (*moral identity* module and *coherence of science* module; 1: present, -1: absent), and 3) the two-way interaction between the first two factors, while controlling for the same covariates (see Figure 1.4).

The *climate science* module showed opposite effects on the probability of selecting an additional political climate action depending on the presence or absence of the complementary coherence questions. The combination of the *moral identity* module and the *coherence-of-science* module itself did not significantly change the odds of selecting an additional political climate action relative to the control condition, $OR = 1.15, p = .73$. With the *climate science* module present, however, the combination of the other two modules more than quadrupled the odds of selecting one more political action, $OR = 4.31, p < .001$. The *climate science* module alone had a negative effect, $OR = 0.40, p = .032$, in that the presence of the climate question decreased the odds of selecting an additional political climate action relative to the control. In the presence of the other two modules, the effect of the climate science module turned positive, but the module did not significantly raise that odds, $OR = 1.50, p = .30$. The interaction between the two factors was significant, $OR = 3.74, p = .024$. These results suggest that the combination of the *modal identity*

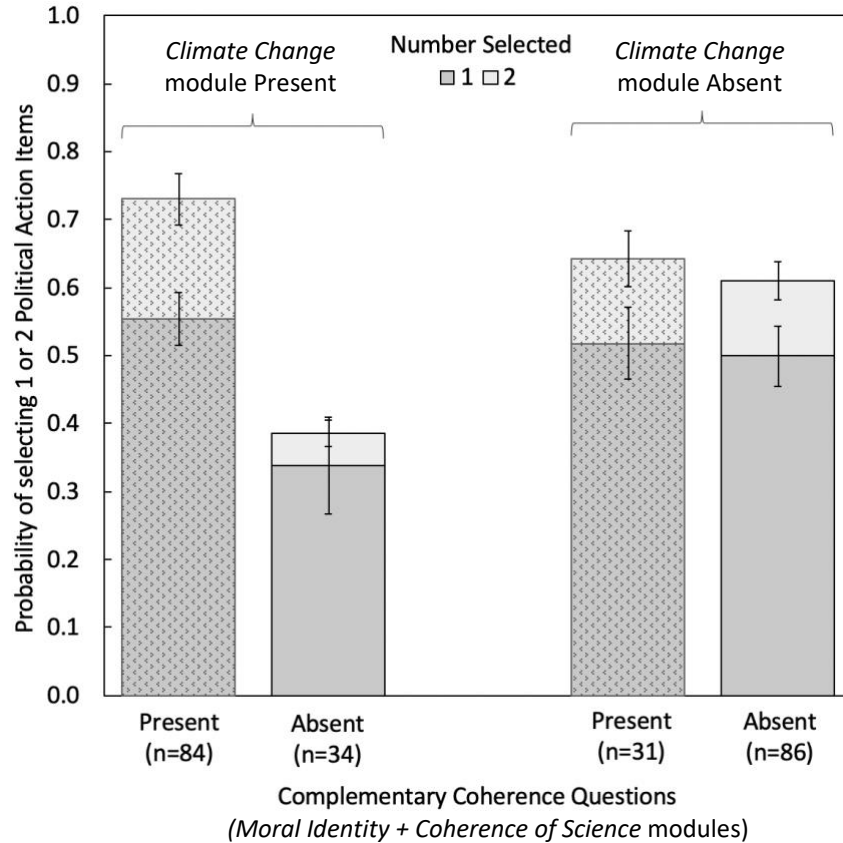


Figure 1.4. Probabilities of selecting zero, one, or two political climate actions, estimated by the presence/absence of the *climate change* module and the presence/absence of the other two modules (four bars from left to right correspond to conditions a, e, b, and f from Table 1, respectively). ‘Complementary Coherence Questions’ here denote the combination of the *moral identity* module and the *coherence of science* module. The unmarked height above each bar indicates the estimated probability of selecting zero action.

module and the *coherence-of-science* module is more effective in promoting climate actions than direct presentation of scientific evidence for ACC.

The control participants’ probability of selecting a climate action may seem surprisingly high. A probable cause is the positive wording of our climate action instructions, which stated

(uniformly across conditions), “As individuals, we each can take impactful actions to protect and support the flourishing of our planet. To that end, what actions are you willing to take?” Highlighting positive actions that individuals can take to promote a better society has been found to increase intention to take pro-climate actions (Bain et al., 2012). This positive effect may be expected to superpose on the effects of our interventions.

Liberals and conservatives did not significantly differ in the interaction between climate and complementary coherence, $OR = 1.37$, $p = .79$. This result was obtained by a three-way factorial ordinal logistic regression analogous to that for the interaction between the *moral identity* module and the complementary coherence from the first analysis. Compared to the climate-only condition, adding the complementary-coherence questions raised the probability of selecting political climate actions for both conservatives and liberals, significantly so for liberals, $p < .001$, although not so for conservatives, $p = .42$ [for this analysis, the sample size happens to be very small for conservatives who had low-science-fact knowledge ($n = 3$), the subgroup for whom the benefit of the complementary-coherence questions was the greatest].

Overall, inducing participants to adopt a morally responsible, pro-scientific, pro-environmental worldview seemed to have a significant effect on their willingness to take political pro-climate actions.

Discussion

The current study recruited participants from the ten U.S. states with the highest level of climate skepticism and found that conservatives’ willingness to take climate actions decreased with increasing levels of science knowledge, consistent with previous reports (Drummond & Fischhoff, 2017; Funk et al., 2019; Kahan et al., 2012). In the polarized milieu in the U.S., efforts

to promote climate actions by conveying climate science across broad sectors of “global warming’s six Americas” (Leiserowitz et al., 2018) may seem futile. But our study demonstrates otherwise.

These findings show that climate-change persuasion can be effective when we use educational materials that leverage two basic cognitive constraints—causal invariance and coherence. The most effective materials were those that went beyond conveying factual climate-science information to encourage the construction of a coherent personal moral narrative consistent with general science and climate science. Specifically, for both conservatives and liberals, embedding ACC information in materials on 1) incontrovertible everyday observations for which science gives explanations that are recognized as more coherent than a reasoner’s own, and 2) victims suffering from a weather catastrophe that can be coherently explained by ACC, raised willingness to take climate actions in comparison to presenting ACC information alone.

Our intervention was most effective when all three components were jointly presented to create a moral and rational narrative. The effect of our full-coherence intervention cannot be explained away by any of the three individual components in the intervention: ordinal logistic regressions showed that presenting each of the three components alone resulted in a decrease in willingness scores compared to the control condition. Therefore, our full intervention cannot simply be reduced to priming an empathetic mindset using a moral identity module (Batson et al., 1997; Petty & Cacioppo, 1986), or evoking concrete personal experiences as support for our arguments using the coherence-of-science module (Kubin et al., 2021),

In particular, the moral identity module was positioned at the beginning of our full intervention with the goal of clearly setting the general moral stance of our experiment. A similar opening instruction was used by Ranney and Clark (2016) prior to showing educational materials explaining the mechanism of global warming to participants. The researchers assured participants

that all materials are true, they are welcome to check the validity of the materials on the cited websites, and they are encouraged to share the information from the study with their family and friends. These instructions did not involve any concrete personal experience. Yet, they appeared to have convinced participants to believe that the intent of the study is to distribute scientific knowledge about climate change to the public so that they can make more rational and informed decisions. Although subtle, these instructions and our moral identity module may bolster participants' self-image of a rational and caring American citizen, produce a moral common ground between researchers, scientists, and participants, and strengthen the credibility of the researchers and the provided materials.¹

Our findings also show that while resonance with the coherence of scientific explanations raises willingness to take pro-climate actions, knowledge of science- and climate-facts does not. Science facts and skills are tools for serving specific goals (Drummond & Fischhoff, 2017; Kahan et al., 2012), rather than general constraints essential to cognition itself, constraints that form the human mind's natural immune defense against maladaptive beliefs. The fact that brief science exercises (taking less than 25 min on average) designed for the current study had significant effects suggests that current science education in the ten U.S. states where we conducted our study, but possibly more generally, does not often convey the parsimony of explanations in a relatable form, and that basic cognitive constraints can guide effective science education. They can also guide rational decisions when trustworthiness of information sources alone does not.

¹ Study 1 varied the position of the moral-identity module, to be either at the beginning or at the end of the intervention materials. The two positions had a similar effect on the dependent measures.

The universal need for a coherent narrative (Ochs & Capps, 2001) paradoxically also governs conspiracy theories denying ACC, despite their incoherence from a scientific perspective (Lewandowsky et al., 2018). Kahan et al. (2017) found that when someone's worldview, identity, or livelihood are threatened by climate or environmental mitigation policies, they frequently engage in "identity-protective cognition". Lewandowsky et al. examined conspiracy theories and found that they characteristically involve logical contradictions at the level of specific propositions about climate. The authors note that coherence is achieved at an abstract level, 'namely that "something must be wrong" with the scientific evidence in order to justify a political position against climate change mitigation' (p. 175). The denial of scientific evidence and the mistrust of climate scientists may be viewed at a more abstract level to be a "simulation of coherence by conspiracies" (p. 175) to achieve a narrative of a moral self.

The disinformation is reified as facts in echo chambers (Mitchell et al., 2014; Rosentiel, 2009). Conformity pressure and the process of confidence evaluation independently and jointly result in majority views being held with greater confidence, regardless of whether the views are accurate (Koriat et al., 2018). Across a wide range of stimuli (e.g., which of two shapes has a larger area? is avocado a fruit?), psychological studies show that people's subjective confidence on a judgment uses the consistency of judgments and coherence with other beliefs as "proxies" for truth (Koriat, 2012, 2018; Koriat & Adiv, 2016). The findings on subjective confidence follow from the under-determination problem: no matter how much we know about the world, "we have no knowledge about the world over and above what we know about it" (Koriat, 2018, p. 620). Our representation of reality is our only conception of it. Conspiracy theorists exploit this basic fact underlying cognition to fabricate their lies. Thus, despite the logical contradictions in conspiracy theories, the theories are coherent and "authentic" to adherents within their echo chambers.

This study harnessed the same need for coherence to reinforce laypeople’s appreciation of science by evoking their intuitive resonance with the parsimony of scientific explanations for puzzling everyday events. Ochs and Capps observed that there exists a dialectic tension between the two narrative proclivities to resolve the discrepancies. This tension is likely to be the window where belief change can occur. The current study did not measure conspiracy ideation directly. However, if coherence and causal invariance are constraints that shape our understandings, then the logical consistency of science ought to be capable of trumping the logical inconsistencies of conspiracy theories when those constraints are leveraged.

The cognitive-science framework of our study offers an explanation of *why* the two universal narrative proclivities identified by anthropologists *are* universal (Ochs & Capps, 2001): Central to the functioning of human cognition is the need to carry a coherent narrative of our lives—an adaptive solution to the problem of under-determination. Recognizing this need may enable us to harness its power to change minds toward more rational actions.

Part 2: Effect of Immoral Character on Non-Moral Judgments

The rise of the #MeToo movement in recent years has brought revelations of many cases of sexual violence committed by people in powerful positions. Much to the public’s shock and dismay, some of the figures accused by alleged victims were not only respected for their expertise, skill, and talent, but also for their moral character and likability. Bill Cosby, for example, was accused by more than 50 victims of sexual assault that spanned decades (Francescani & Fisher, 2019). His case was particularly appalling to many because of the wholesome fatherly persona he had presented to his audience.

History has seen many cases in which highly successful figures have been of dubious moral character. The conflicting information about those individuals raises one of the long-standing questions—whether art can be separated from the artist. Can we denounce Richard Wagner for his strong anti-Semitic beliefs while still enjoying his opera? Can we make a fair judgment of Pablo Picasso’s brilliance as an artist after realizing that he was abusive to multiple women and his son? In this part, we focus on how discovery of immoral acts ‘taints’ judgments of the past accomplishments and competence of individuals who are prominent in their field—a situation increasingly common in the age of internet and new media.

Judgments about other people often incorporate a variety of information with conflicting valences. The interplay of positive and negative information about a target individual can result in complex patterns of social judgments and decisions (Ditto et al., 2009; Holyoak & Simon, 1999; Kunda & Thagard, 1996; Simon et al., 2015). Imagine that a kind and likable person is ineffective at work. Her colleagues’ critical judgments about her competence might be mitigated because of her likable personality. Even worse, if her perceived lack of competence is extreme, that information might negatively affect colleagues’ judgment of her personality and moral character.

As reviewed in the introduction, a longstanding view of impression and attitude formation postulates that the process is driven by cognitive and motivational pressure for coherence (Asch, 1946; Heider, 1946). A coherent view of others is achieved by shifting inconsistent perceptions and beliefs so as to increase their coherence with other information (Ditto et al., 2009; Simon et al., 2015). An important example of this line of work involves the “halo effect” (e.g., Nisbett & Wilson, 1977). Researchers have found that the global impression of an individual (e.g., friendly vs. distant) alters evaluations of more specific attributes, such as likability and attractiveness of the individual’s appearance.

But what are the primary types of information that people use in social perception? Competence and warmth are two basic dimensions that have been considered fundamental to impression formation (Abele & Wojciszke, 2007; Fiske et al., 2007; Judd et al., 2005; Wojciszke et al., 1998). The first dimension (competence/agency) has typically been linked to an individual's ability, intelligence, and skill, whereas the second dimension (warmth/morality/communion) has often been linked to likability, friendliness, and trustworthiness in social relationships. Wojciszke et al. (1998) found that 82% of the variability in undergraduate students' global impressions of others was explained by these two basic dimensions.

The intrinsic motivation to seek potential cooperators who are generous and caring leads people to consider traits that signal warmth, sociability, and morality as generally more important than traits linked to competence (Goodwin, 2015; Wojciszke et al., 1998). Brambilla and colleagues (2019) showed that morality information has primacy over other kinds of information in impression updating. Following Goodwin (2015; Goodwin et al., 2014), they divided the warmth dimension into sociability (friendliness, likability, and kindness) and morality (honesty, sincerity, and trustworthiness). Their participants were given two descriptions of behaviors of a target person that differed in valence (positive vs. negative) and dimension (morality vs. sociability vs. competence). All descriptions were matched on the absolute value of the valence ratings. Participants rated the global impression of the target (on a scale from extremely negative to extremely positive) each time they received a description of behavior. Results showed that information on the morality dimension had a greater effect on impression updating than did the other dimensions.

Based on these findings as well as the coherence-based reasoning framework, I hypothesize that non-moral judgments about an individual's competence, accomplishments, and contributions

can be influenced by moral evaluations of her character, even in cases where the two are technically orthogonal to each other. More specifically, this influence of immoral character on non-moral judgments is expected to be channeled through the decrease in the target's perceived likability or relatability. That is, the decrease in general likability and willingness to relate to the target person may lead people to discount their non-moral evaluations of the target as well.

Study 2 focuses on laypeople's evaluations of successful figures who were 'disgraced'—that is, those who committed serious moral transgressions. Such evaluations require resolution of conflicts between strongly positive and negative information, which can be explained from the perspective of coherence-based reasoning. An online experiment was conducted to measure participants' non-moral judgments about fictitious professionals in various areas. Their ratings were compared between conditions where the professionals did or did not commit a serious moral transgression.

Study 2

The current study resembles the experiments conducted by Stellar and Willer (2018), who focused on the interplay between ratings of morality and other dimensions. Their participants were presented with information about a target person that signaled their moral character (e.g., 'stole expensive items from a store', 'diligently cared for a parent'), and level of competence ('He has been working at his company for five years and has done a reasonably good job') in varying formats and contexts. Stellar and Willer's participants then rated the target individual on the competence dimension (e.g., 'how good do you believe he is at his job?'). In both within- and between-subject designs, they found that immoral character significantly lowered competence ratings. This pattern contrasted with findings from another survey in the same paper where

participants responded that evaluation of moral character should be made independently of competence judgments.

In Study 2, I expanded on Stellar and Willer's (2018) findings using more realistic settings and a richer set of descriptions about targets. In the experiments by Stellar and Willer, the target individuals were mostly ordinary people without any notable accomplishments, and only minimal descriptions were provided. Thus, the strong influence of information about immorality observed in their studies may have been due to the fact that their participants were given little evidence to support the competence of the targets.

In the present study, participants were shown vignettes that described fictitious successful professionals working in several different fields. The descriptions included detailed information related to their professional competence, credentials, and achievements. By providing realistic details in the descriptions, participants' social judgments can be based on richer information, as would be the case if they were making similar judgments about their friends, acquaintances, or influential public figures.

Stellar and Willer (2018) further proposed that perceived social intelligence—the ability to understand and deal with others and to know social rules (Kosmitzki & John, 1993)—of the target person serves as a mediator variable linking morality and competence judgments. That is, they suggested that participants perceive targets with immoral character as having lower social intelligence, which leads to lower judgment of competence, especially if a relevant task requires social interaction. In addition to social intelligence, the current study proposes psychological involvement/distance with the target as another potential mediator. Given our general motivation to maintain a positive self-image (e.g., Alicke & Sedikides, 2009), it was predicted that immoral character will lead participants to feel immediately detached from the target individual, thereby

decreasing their non-moral evaluations of the target. The influence of psychological involvement may be especially potent when participants are given rich and detailed information about the target. Previous studies on the impact of immorality (e.g., Brambilla et al., 2019; Stellar & Willer, 2018) did not provide detailed information about the target, and did not measure psychological involvement with targets.

As a secondary hypothesis, I posited that the decrease in non-moral judgments due to immoral character will be greater in entertainers (e.g., actor, musician) than in researchers (e.g., biologist, philosopher). If our person evaluation is more dependent on the likability and personal qualities of the targets when evaluating entertainers compared to when evaluating researchers, the effect of moral violation may be greater for entertainers.

Method

Participants and Design

A priori power analysis was conducted using G*Power. In order to detect a small-to-medium ($\eta_p^2 = .02$) effect with a power of .95, α level of .05, and a correlation of .4 between repeated measurements (estimated from a pilot study) in a 2×2 mixed ANOVA design, 194 participants would be needed. Amazon MTurk was used to recruit 230 participants residing in the U.S. Of these participants, 28 were dropped from the analysis because they did not pass either of two attention checks, leaving 202 participants (93 females, $M_{age} = 38.2$, $SD_{age} = 11.6$) They received \$1.0 as compensation for their participation. Median time spent on the entire survey was 5.3 minutes.

The experiment had a 2 (profession type: entertainer/researcher) \times 2 (moral violation: present/absent) design. Profession type was manipulated between subjects, whereas the moral violation condition was manipulated within subjects.

Materials and Procedure

Participants were instructed to read descriptions of two professionals in different fields on a Qualtrics survey and make judgements about them after careful consideration. Each description presented a hypothetical professional, complete with an AI-generated face of a middle-aged Caucasian male (from <https://thispersondoesnotexist.com>; two of the nine preselected images were randomly shown to each participant as two professionals). A brief biographical description was provided, with an overview of the target's main accomplishment in their respective field. The last paragraph of the description for the control condition detailed the professional's hobby (either creative writing or outdoor activities), whereas the analogous paragraph for the violation condition described a seriously immoral action (either multiple sexual assaults or a drug-related murder). All four manipulation paragraphs were matched in length (about 75 words). For example, one of the introductions describing a biologist in the violation condition was the following:

Raymond Johnson is a biologist who specializes in ichthyology. More specifically, his main research interest is in the evolution of ancient species of fish. He received his PhD from the West Virginia University in 1990, and is now an associate professor at Davidson College. Johnson received the Annual Excellence in Research Award from Davidson College in 2011. Below we describe one of his most important discoveries.

Main Accomplishment: Investigation of Evolution of Coelacanth Fish

Johnson studied the evolutionary timeline that shows how modern tetrapods and lungfish can be traced back to coelacanth fish. In his most well-known paper, Johnson presented convincing evidence that three-lobed tails of coelacanth fish are used to swim. This finding helped ascertain the fact that modern-day fish are related to lungfish and tetrapods.

In 2018, Johnson's life took a dramatic turn. His wife stood up to reveal to the public that she had been physically abused by Johnson for many years. Even worse, soon after the investigation on his domestic violence began, four other women told the media that they were sexually harassed and assaulted by Johnson while he was traveling. After carefully considering multiple pieces of evidence and witnesses, the investigators concluded that the allegations were all true.

Participants were then asked to make social judgments about the professional on four separate axes: competence (two items, $r_s > .54^2$; *Andy Goodwin is a talented actor*³; *Andy Goodwin is a brilliant actor*), accomplishment (three items, $\alpha_s > .83$; e.g., *His album "Beyond the Sound" has likely progressed music in concrete ways*), contribution (three items, $\alpha_s > .79$; e.g., *Society has benefited from Simon Graham*), and psychological involvement with target (four items, $\alpha_s > .81$; e.g., *I feel that I can relate to Robert Wilson*; *I see Robert Wilson as a natural, down to earth person*; *I am interested in Robert Wilson*; *When met in person, I think Robert Wilson will make me feel comfortable, as if I am with friends*). Three of the four involvement items were modified from Bocarnea and Brown's (2007) Celebrity-Persona Parasocial Interaction Scale. Participants rated the degree to which they agreed with these statements on a 7-point Likert scale (1: *strongly disagree*, 2: *disagree*, 3: *slightly disagree*, 4: *neither agree nor disagree*, 5: *slightly agree*, 6: *agree*, 7: *strongly agree*). These survey items were tailored to the description of each professional and accomplishment being evaluated. For example, in one of the accomplishment questions, the blanks

² All reliability measurements were calculated separately for the violation and control conditions. The lower statistics are reported here.

³ To provide a wide set of examples used in the experiment, the questions shown here are taken from materials created for different professionals with randomly assigned names.

in the statement “[Main accomplishment] has likely progressed [his field] in concrete ways” were filled differently for each professional.

An additional set of questions was presented to measure the perceived social intelligence of the professional (Kosmitzki & John, 1993), in order to replicate Stellar and Willer’s (2018) findings. For this measurement, participants read 10 attributes (e.g., Knowing social rules and norms; Understanding people; Open to experiences and ideas), and reported the extent to which they thought each attribute would be characteristic of the professional ($\alpha s > .88$). The social intelligence items were the same across all conditions and were rated on a 5-point Likert scale (1: very uncharacteristic, 2: uncharacteristic, 3: uncertain, 4: characteristic, 5: very characteristic).

The two fictitious targets shown to each participant had the same type of profession: entertainer or researcher. Participants in the entertainer group were shown descriptions of either a comedian and an actor or a musician and an athlete, whereas participants in the researcher group were shown descriptions of either an economist and a philosopher or a chemist and a biologist. The two descriptions in each of the pairs were matched in their levels of perceived accomplishment, competence, and contribution ratings as much as possible based on the results from an independent pilot study ($n = 29$). The pairs of professionals, the order of presentation of the two professionals (e.g., musician \rightarrow athlete or athlete \rightarrow musician), the order of violation and control manipulations (violation \rightarrow control or control \rightarrow violation), and the type of violation (sexual assaults or drug-related murder) were all independently counterbalanced within each of the entertainer and researcher groups. The type of filler paragraph in the control condition (creative writing or outdoor activities) was randomly determined for each participant by Javascript codes embedded in the Qualtrics survey. Attention check questions were embedded in the list of social intelligence

questions, instructing participants to “choose ‘very uncharacteristic’”. Data from participants who did not choose ‘very uncharacteristic’ at least once were excluded from analyses.

After the survey items for the second professional were completed, participants were debriefed and answered questions regarding demographics and believability of scenarios. About 64% of the participants believed that the stories were real, 25% suspected that the stories were not real, and 11% did not believe that the stories were real. Hence, the scenarios were perceived by many of our participants to be realistic.

Results

The pattern of results is summarized in Figure 2.1. A 2 (profession type) \times 2 (morality violation condition) mixed ANOVA for the competence score was conducted. Participants in the control condition rated the professional's competence as higher ($M = 5.70$, $SD = 0.88$) than they did in the moral violation condition ($M = 5.43$, $SD = 1.14$), $F(1, 200) = 10.00$, $p = .002$, $\eta_p^2 = .048$. The two profession types did not differ overall in the rated competence, $F(1, 200) = 0.09$, $p = .92$, and the interaction between the two factors was not significant, $F(1, 200) = 0.03$, $p = .86$. To control for the difference due to profession type, follow-up paired t -tests were conducted using the within-subject difference of competence scores (control – violation) for entertainer and researcher groups separately. Both entertainers ($t(103) = 2.12$, $p = .036$, $d = 0.21$) and researchers ($t(97) = 2.36$, $p = .021$, $d = 0.24$) showed a significant decrease rated competence ratings because of the moral violation information.

An analogous ANOVA for the accomplishment score was conducted. Accomplishment scores were significantly lower in the moral violation condition ($M = 4.96$, $SD = 1.40$) than in the control condition ($M = 5.38$, $SD = 0.99$), $F(1, 200) = 20.16$, $p < .001$, $\eta_p^2 = .092$. Scores were lower overall for entertainers than for researchers, $F(1, 200) = 15.46$, $p < .001$, $\eta_p^2 = .072$. The interaction

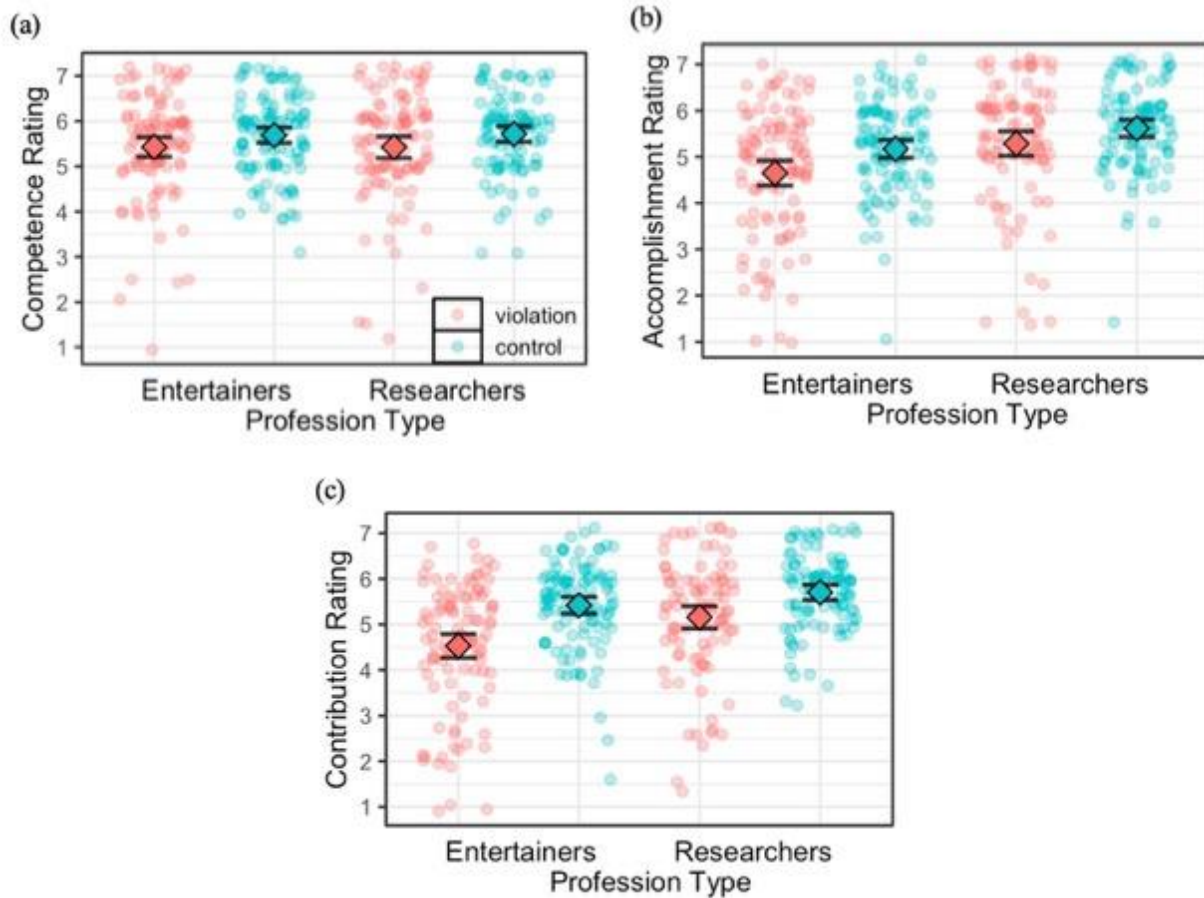


Figure 2.1. Ratings of (a) competence, (b) accomplishment, and (c) contribution in the presence versus absence of information about a moral violation by entertainers (left two columns in each panel) and researchers (right two columns in each panel). Diamonds indicate condition means. Error bars indicate 95% confidence interval.

was not significant, $F(1, 200) = 1.07, p = .30$). In the follow-up paired t -tests, both entertainers ($t(103) = 4.06, p < .001, d = 0.40$) and researchers ($t(97) = 2.35, p = .021, d = 0.24$) showed a significant decrease in rated accomplishment when described as guilty of immoral actions.

Finally, an ANOVA for the contribution score was conducted. Contribution scores were significantly lower in the moral violation condition ($M = 4.83, SD = 1.34$) than in the control

condition ($M = 5.56$, $SD = 0.91$), $F(1, 200) = 57.57$, $p < .001$, $\eta_p^2 = .224$. Scores were lower overall for entertainers than for researchers, $F(1, 200) = 12.98$, $p < .001$, $\eta_p^2 = .061$. The interaction was marginally significant, $F(1, 200) = 3.36$, $p = .068$, $\eta_p^2 = .017$. Follow-up paired t -tests were performed using the within-subject difference scores as the dependent variable (control – violation). Both entertainers ($t(103) = 6.63$, $p < .001$, $d = 0.65$) and researchers ($t(97) = 4.10$, $p < .001$, $d = 0.41$) showed a significant decrease in rated contribution when described as guilty of immoral actions.

Overall, results suggested that the negative influence of moral violations on the dependent measurements was not significantly different between entertainers and researchers. Although the effect size of moral violation seemed to be somewhat larger in the entertainer condition than in the researcher condition, mean comparison results were not conclusive.

Mediation Analysis.

To conduct mediation analyses, I used MEMORE version 2.1, a macro for SPSS developed to conduct mediation and moderation analyses with repeated measures (Montoya & Hayes, 2017). All mediation analyses reported here used $n = 20,000$ as sample size for bootstrapping.

First, I attempted to replicate Stellar and Willer's (2018) finding that perceived social intelligence mediates the influence of morality on competence judgment. After collapsing across the entertainer and researcher conditions, social intelligence rating was entered as the only mediator variable ('model=1' in MEMORE). Results showed that the direct effect of immorality on competence judgment was not significant ($b = -0.07$, 95% CI = [-0.30, 0.16], $t(199) = -0.59$, $p = .55$), whereas the indirect effect through social intelligence was significant ($b = -0.20$, 95% CI

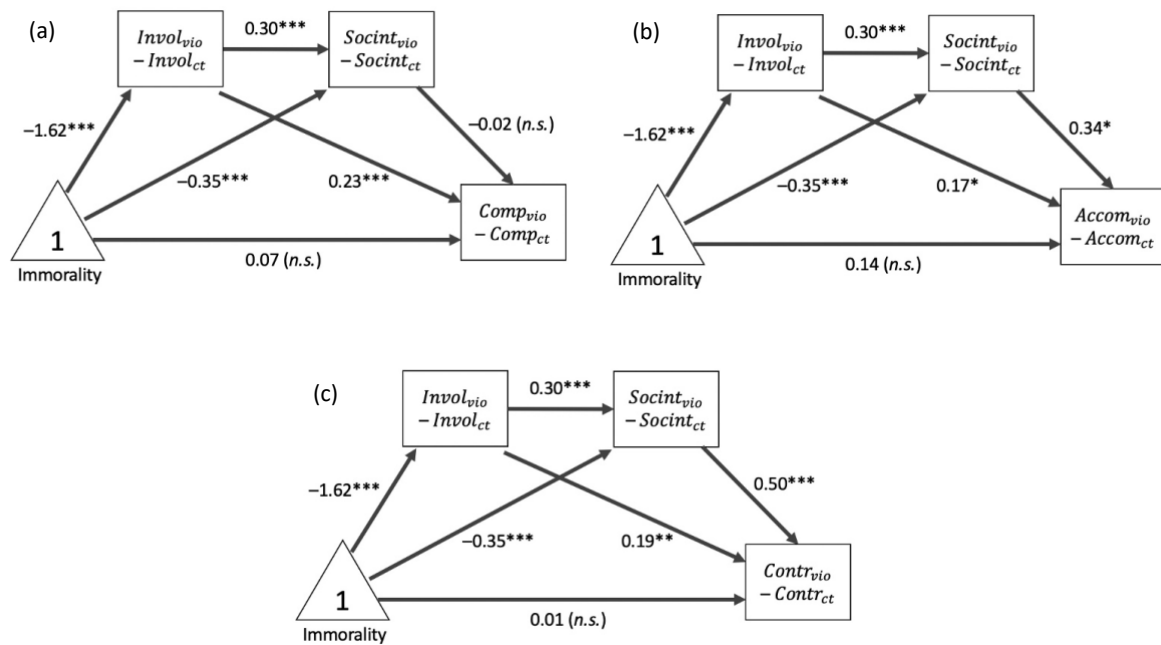


Figure 2.2. The effect of immorality on ratings of targets' (a) competence (*Comp*), (b) accomplishment (*Accomp*), and (c) contribution (*Contr*) is mediated by involvement with target (*Invol*) and/or perceived social intelligence of target (*Socint*). Weights indicate unstandardized regression slopes. MEMORE uses the within-subject difference of each of the variables as nodes in the model. Accordingly, the weights of arrows from the intercept (lower left triangle in each pane) indicate the within-subject effects of violation. Residuals and centering terms are omitted for clarity. $*p < .05$, $**p < .01$, $***p < .001$.

= [-0.43, 0.00]). Next, a more comprehensive mediation model including both involvement with target and perceived social intelligence as parallel mediators was tested ('model=1' in MEMORE). After being distinguished from the indirect effects, the direct effect of immorality was not significant ($b = 0.07$, 95% CI = [-0.17, 0.32], $t(197) = 0.60$, $p = .55$). However, the indirect effect of immorality through social intelligence was also not significant ($b = 0.02$, 95% CI = [-0.31, 0.31]), whereas the indirect effect through involvement with target was significant ($b = -0.37$, 95%

CI = [-0.62, -0.12]). This result indicates that participants' psychological involvement with the target was a stronger mediator than social intelligence in explaining the impact of moral taint. The components of 'hot cognition'—in this case, likability or motivation to relate to the target—were important connectors between moral and non-moral evaluations.

Next, serial mediation models were tested in which both mediators were entered and the first mediator also predicted the second mediator (see Figure 2.2; 'serial = 1' in MEMORE). I hypothesized that immorality would have an immediate negative effect on involvement with target, which will then decrease perceived social intelligence. Then, all three variables were tested as predictors for the competence rating (Figure 2.2). For simplicity, I will refer to the indirect effect from immorality to involvement to competence as *ind1*. All links along *ind1* were significant. Neither the indirect effect from immorality to social intelligence to competence (*ind2*; $b = 0.01$, 95% CI = [-0.13, 0.12]), nor that from immorality to involvement to social intelligence to competence (*ind3*; $b = 0.01$, 95% CI = [-0.18, 0.18]) was significant. Thus, the influence of immorality on competence judgments was only channeled through a decrease in participants' psychological distance with the target, rather than through the perceived social intelligence of the target (see Figure 2.2a).

The same serial model was also used to predict accomplishment and contribution ratings (Figure 2.2b and 2.2c). When predicting the accomplishment rating, the direct effect of immorality was not significant ($b = 0.14$, 95% CI = [-0.11, 0.39], $t(197) = 1.11$, $p = .27$), whereas *ind1* ($b = -0.28$, 95% CI = [-0.52, -0.03]), *ind2* ($b = -0.12$, 95% CI = [-0.27, -0.01]), and *ind3* ($b = -0.57$, 95% CI = [-0.83, -0.34]) were all significant. When predicting the contribution rating, the direct influence of immorality was not significant ($b = 0.01$, 95% CI = [-0.21, 0.24], $t(197) = 0.10$, $p = .92$), whereas *ind1* ($b = -0.31$, 95% CI = [-0.57, -0.07]), *ind2* ($b = -0.18$, 95% CI = [-0.31, -

0.07]), and *ind3* ($b = -0.25$, 95% CI = [-0.43, -0.10]) were all significant. In sum, participants' devaluations of the past work and contributions of immoral professionals were mediated through multiple pathways involving psychological distance with the target and perception of the target's social intelligence.

Discussion

Forming evaluations of others requires incorporating a rich and diverse set of information, which often involves conflicting valences. I propose that a coherence-generating mechanism underlies the process in which evaluations of other people are formed. In particular, negatively-valenced information about moral character was expected to trigger negative shifts in evaluations of professional competence and achievements, thereby increasing overall coherence of the impression of the target.

The current study experimentally tested this prediction by examining how people's evaluations of a target's competence, level of accomplishment, and the magnitude of their contributions to their field shift to cohere with knowledge of the target's immoral actions (sexual assault or murder). Replicating previous work, we found that evaluations were significantly reduced when people learn of a moral transgression, indicating that evaluations of professional attributes shift to cohere with moral-based evaluations of their behaviors. Notably, the increased negativity of these professional evaluations was found both for a continuing trait and for past activities (accomplishments and professional contributions in the past).

Mediation analyses provided a more detailed picture of the mechanisms by which information about immorality impacts other judgments. I assessed the role of perceived social intelligence (identified as a mediator by Stellar & Willer, 2018), and also subjective involvement with the target (including perceived likability of target). For competence judgments, the pathway

through involvement was the only reliable predictor; for judgments of accomplishment and contribution, perceived social intelligence acted as an additional mediator. For all three judgments, the direct path from immorality was not significant after accounting for the mediator variables.

The impact of involvement as a mediator suggests that the observed coherence shifts are motivated by the desire to maintain a positive self-image (Alicke & Sedikides, 2009), which leads people to psychologically distance themselves from a target described as immoral. Other mediators may certainly exist, including affective (e.g., anger toward the target) and non-affective (factual inferences about target's career or the reported transgression) variables (cf. Simon et al., 2015). Future work is needed to more closely examine the role played by moral considerations in forming judgments of the achievements of professionals. It would be useful to manipulate the severity of the moral violation (e.g., tax evasion versus murder) to assess which types of moral transgressions induce the largest shifts in judgments. It would also be useful to examine the impact of moral violations that are more directly pertinent to professional situations (e.g., plagiarism by a journalist).

The #MeToo movement on Twitter has brought a great many instances of sexual abuse to the spotlight. Much of the discourse has focused on alleged abuse by well-known professionals and celebrities. As a follow-up study, text-based analyses of social media data (e.g., Twitter conversations about the #MeToo movement) could illuminate how evaluations of professional achievements shift in the aftermath of coverage of immoral acts, and differences related to the perceived nature of those acts. For example, it may be possible to assess which moral foundations are activated when disgraced professionals are discussed by performing computational analyses using the eMFD software (Hopp et al., 2020), a toolkit for extracting moral sentiment from natural

language text. Such research could help to reach a richer understanding of how moral evaluations trigger reevaluation of contributions made by professionals to their field and the world.

Part 3: Collective Guilt, Pride, and Responsibility

Dynamics within and between groups is one of the substantive factors that guide our moral judgments and decisions. Relationships between members of a social group can produce special rights, obligations and *agent-relative* responsibilities that people impose on themselves and others (Nagel, 1986; Lee & Holyoak, 2020). For instance, a parent feels obligated to provide food and care for her child more than for someone else's child. Note that these responsibilities emerge from moral principles that hold for particular relationships and groups (e.g., 'Parents have a duty to provide care to their children') rather than from the involved individuals' specific conduct in the past (Holyoak & Powell, 2016).

The assignment of moral responsibility to groups can be viewed as an extreme example of this phenomenon on a broader scale; That is, simply being a member of the same group with a wrongdoer sometimes leads people to feel guilty and obligated to make reparations for the harm done (Ferguson & Branscombe, 2014). Furthermore, to the extent that people experience such a shared responsibility within their groups, they may also be inclined to impose analogous responsibilities to individuals in other groups based on group affiliation. Indeed, studies across cultures and disciplines have repeatedly documented the existence of collective moral responsibility: over several generations, White Americans have expressed guilt for the enslavement and mistreatment of African Americans and indigenous peoples by early White Americans (Chudy et al., 2019; Iyer et al., 2003; Swim & Miller, 1999); postwar generations in

Germany have been apologetic to Jews for the atrocities committed under the rule of the Nazi party (Wohl et al., 2006); in an ethnographic study that analyzed the electronic Human Relations Area Files (eHRAF; Curtin et al., 2020) indications of collective guilt were found in 45 out of 71 societies around the world. Jaspers (2009) distinguished between guilt based on one's own conduct and guilt based on one's identity, dubbing the latter "metaphysical guilt"; others have termed it "moral taint" (Silver, 2006; Uhlmann et al., 2012).

Yet, the idea of assigning moral responsibility or blame to individuals who did not causally contribute to the wrongdoing is unusual in most of the discussions in Western philosophy (Smiley, 2017). Philosophers have postulated that, in a typical setting, to hold an individual responsible or blameworthy for a wrongdoing requires assuming that the individual has acted as a moral agent. That is, the individual must have acted freely rather than under coercion to cause harm, intended the action, and believed or knew that their action was wrong (for a philosophical review see Talbert, 2019). Hence, blaming people based on their personal ties to a wrongdoer is typically dismissed as the fallacy of guilt by association. In line with this view, many studies in moral psychology focused on confirming that the mental states of the agent inferred by the reasoner—especially those pertaining to whether the agent had control over their action and harm—are indeed critical determinants of moral judgments of responsibility, wrongness, and blame (e.g., Alicke, 2000; Cushman, 2008; Cushman & Young, 2009; Greene et al., 2009; Lagnado & Channon, 2008; Malle et al., 2014; Giffin & Lombrozo, 2018; for reviews see Bartels et al., 2015; Guglielmo, 2015).

How can this gap between psychological theory and reality be closed? One way to do so is to acknowledge the importance and prevalence of agent-relative rights, rules, and obligations in our world (Holyoak & Powell, 2016; Nagel, 1986). That is, descriptive theories of morality may need to refrain from expecting a few context-independent, agent-neutral moral principles (e.g.,

categorical imperatives, utility maximization) to determine our judgments and decisions under a wide variety of situations. On the other hand, models of everyday moral reasoning should attempt to identify and incorporate more specific factors such as cultural/social norms and deontological rules (e.g., rights, and duties) that must be honored in particular circumstances.

In the context of this dissertation, assignment of responsibility in the absence of causal contribution exemplifies the significance of non-logical relations between mental constructs, including moral judgments, moral obligations, and internalized social norms. Especially when an agent did not benefit from a past wrongdoing by a member of the same group and the shared group membership is tangential to the wrongdoing, there is no rational basis to impose responsibility to make reparations to the agent. However, the widespread effect of shared group membership on responsibility judgment suggests that laypeople's moral judgments are often guided by cognitive consistency between the emotional valences tied to objects and people (e.g., likability of, and sympathy toward wrongdoer, agent, and victim; Kunda & Thagard, 1996; Simon et al., 2015), moral obligations (e.g., reparation must be made for a past harm; Holyoak & Powell, 2016), and social identity (e.g., how much the actor and wrongdoer identify with the affiliated group).

Holyoak and Powell (2016) conceptualize rights and duties as interdefined pragmatic reasoning schemas (Hohfeld, 1919; Cheng & Holyoak, 1985; Holyoak & Cheng, 1995) that operate between moral agents under specific conditions. For instance, the right of agent *X* against agent *Y* with respect to action *A* corresponds to the duty of *Y* toward *X* with respect to *A*. Applying this abstract schema to the case of collective responsibility assignment, a conditional agent-relative moral rule can be postulated: if one group has inflicted a serious harm on another group in the past, the members of the wrongdoer group have a special obligation to make reparations to members of the victim group. Conversely, from the perspective of the victim group, the same rule can be

viewed as specifying their rights to receive appropriate reparations from the wrongdoer group. Then, the acceptance of such rule or injunctive social norm results in the activation of motivations and goals (e.g., motivation to see the victim group get compensated by the wrongdoer group) during reasoning, which can be modeled as a constraint satisfaction process (Simon et al., 2015).

One way to examine the influence of such motivations and goals on collective moral judgments is comparing cultures that adhere to different social norms. The degree to which people emphasize group-level moral responsibilities and rules is expected to vary systematically across cultures. In particular, studies have noted the differences in reasoning between individualistic and collectivistic cultures (Oyserman et al., 2002; Triandis, 2001). Typically, Western societies influenced by the Judeo-Christian tradition and the East Asian societies influenced by Confucian tradition have been used to exemplify those two types of cultures, respectively (Bedford & Hwang, 2003; Graham et al., 2016; Henrich et al., 2010; Oyserman et al., 2002; Schulz et al., 2019; Triandis & Gelfand, 1998). Hence, the two studies in Part 3 focused on the cross-cultural differences in the experience and assignment of collective moral responsibility and guilt among group members. It was initially expected that East Asian participants (Koreans and Japanese) will place greater emphasis on the intergroup rights and duties that originate from (historical) group dynamics compared to Americans. In other words, social norms and rules that strongly activate motivations to endorse intergroup duties and rights were expected to increase the assignment of collective responsibility. In addition, the influence of culture was expected to be mediated by how much people perceive the involved groups (their own or someone else's family) as cohesive entities (*entitativity*; Campbell, 1958) and how much they identify with their family in general. The level of cohesion perceived in the wrongdoer group was expected to predict the level of collective responsibility assigned.

In Study 3A, participants were asked to read a scenario involving fictitious characters and groups and make moral judgments from a third-person perspective. In Study 3B, another set of scenarios was written in first-person perspective and participants reported the collective guilt, pride, and responsibility they would experience in the hypothetical situation. Because Study 3B was originally planned to generalize the findings from Study 3A, the results from two studies were expected to converge. However, data showed that whether participants make judgments from a first-person vs. third-person perspective produces a critical difference in how they assign collective guilt, pride, and responsibility. These findings suggest that although culture-/society-specific norms and obligations do have significant effect on people's moral judgments and experiences, the ways in which those effects materialize in concrete cases may not be simple.

Study 3A

This study adapted a paradigm introduced by Uhlmann et al. (2012), who gave participants scenarios that varied the nature of the link between agent (tasked with deciding whether or not to make financial restitution to descendants of the victims) and their grandfather (the wrongdoer). This link was either biological or not (in the latter case, the agent's grandmother had divorced and remarried before he was born, and her second husband was the wrongdoer). Uhlmann et al. argued that people's intuitive belief in essentialism will transfer the moral taint of the older wrongdoer to his descendant more strongly when they were connected through blood ties. The investigators indeed found that participants offered greater compensation when the offender was described as their biological grandfather rather than a grandfather by remarriage.

In order to examine the impact of a wider range of family structures, the present study included a condition involving family connection via adoption, in addition to the remarriage and biological conditions tested by Uhlmann et al. (2012). As a baseline, another control condition was

added where the wrongdoer and the agent were not members of the same family at all. It was hypothesized that perceived “oneness” of the wrongdoer and the agent as members of a common group (*entitativity*) would be highest for the biological condition, followed by the adoptive, remarriage, and no-relation conditions. Acceptance of collective responsibility was assumed to increase with entitativity.

Orthogonally to the relationship conditions, the presence and absence of a financial benefit the agent had received from the wrongdoer was varied. Historical exploitation often results in financial advantages for the exploiter group and disadvantages for the victim group. Radzik (2001; Iyer et al., 2003) have argued that the collective responsibility of members of the exploiter group to make reparations partly derives from the benefits and advantages they inadvertently received. By using a factorial design, the current study could distinguish the effects of financial benefit and relationship on collective responsibility judgment. In addition, I tested whether members of a collectivistic society (Koreans) are more sensitive to the types of relationship between the agent and the wrongdoer than members of an individualistic society (United States). Because collectivistic cultures have often been characterized as valuing interrelations and interdependent aspects of self more than individualistic cultures (Triandis, 2001), it was hypothesized that the effect of type of relationship on collective moral judgments will be stronger in Korea than in the U.S. The individual difference in participants’ identification with their family was hypothesized to mediate the effect of culture.

Method

Participants and Design

The experimental design was a 2 (culture: U.S./South Korea) \times 4 (wrongdoer’s relation to agent: biological grandfather/grandfather in adoptive family/grandfather based on remarriage of

agent's grandmother/no-relation) \times 2 (financial benefit: present/absent) between-subjects factorial design, with each participant reading one scenario. For brevity, the four levels of the relationship variable will be referred to as biological, adoptive, remarriage, and no-relation.

Study 1 of Uhlmann et al., (2012) reported that the effect size of biological (vs. non-biological) family relatedness between grandfather and grandson on recommended reparation was Cohen's d of 0.48. G*Power was used to calculate the sample size required to detect an effect of the same size in pairwise t -tests between four relationship conditions. For a minimum power of .80 and α of .05 in a two-tailed independent samples t -test, 70 participants were required in each group. In order to examine more fine-grained variations in relatedness, 80 participants were recruited in each group.

Amazon Mechanical Turk was used to recruit 649 American participants (283 females, 1 non-binary gender; $M_{age} = 37.8$, $SD_{age} = 11.8$). Around 80 American participants were assigned to each of eight between-subject conditions. American participants were paid \$0.80 for completing the experiment, which took around 4 minutes on average. An additional 358 American participants were recruited but excluded from analyses because they failed to correctly answer either of two comprehension checks.

After collecting data from American participants, 641 Korean participants (347 females; $M_{age} = 36.7$, $SD_{age} = 11.9$) were recruited from Hankook research's (<https://www.hrc.co.kr/eng/>) online survey panel. Quotas were used to roughly match the distribution of age between the American and Korean samples. Participants who failed the comprehension checks were automatically dropped during the experiment in the survey agency's internal system. Korean participants were compensated with points in Hankook research's online system, which were worth around \$0.80. They took about 5 minutes on average to complete the experiment.

Materials and Procedure

The basic content and structure of the scenarios were adapted from Study 1 of Uhlmann et al. (2012). In the American version, all scenarios started by stating the relationship between Sam (wrongdoer), an owner of a small factory in New York, and Brian Johnson, the agent in the scenario. The Korean version was first translated from the English version and then back translated to check for tone and content. In the Korean version of the experiment, the characters were given Korean names and the city was changed from New York to Seoul.

In the no-relation conditions, the scenarios stated that young Brian had been told stories about Sam Miller, a factory owner around his grandfather's age. In the other conditions, how Brian was related to his grandfather Sam was described: either biologically, through adoption, or based on remarriage of Brian's grandmother. The scenarios then described Sam's wrongdoing in the past, which involved exploiting some of the poorest residents of New York as employees. The O'Neal family was the largest group of employees who were especially exploited. Sam threatened them with termination if they complained about their working conditions, even after the two youngest members of the family died while working at the factory.

Next, in the scenario for the no-relation condition with financial benefit present, it was stated that Brian received a scholarship award that was founded by Sam, which made it possible for him to attend college. In the other relationship conditions with financial benefit present, the scenarios stated that Brian inherited some of his grandfather Sam's fortune when he was 20 years old, which made it possible for Brian to attend college. When the no-relation condition was coupled with financial benefit absent, there was no mention of a scholarship. In the other relationship conditions with financial benefit absent, the scenarios stated that Sam's fortune ran out before Brian was born, and Brian received no inheritance from him.

Then, all scenarios described Brian as currently a middle-aged lawyer who recently won the lottery. He decides to donate a portion of his winnings, \$10,000 (converted to South Korean won of roughly equivalent value in the Korean version: 1,000 man-won) to charity, and considers two causes: the International Hungry Children’s Fund, and the education of descendants of the O’Neal family who remain needy today. Participants were asked: “How should he distribute the money between the O’Neal children’s education and the International Hungry Children’s Fund?” Participants could proceed to the next question only when the two responses summed to \$10,000. The amount of money allocated to the O’Neal children’s education was the first dependent measurement.

On the next page, participants were asked: “How responsible do you think Brian should feel for what happened to the O’Neal family?” (5-point scale, 1: not at all responsible, 2: slightly responsible, 3: somewhat responsible, 4: very responsible, 5: fully responsible) This score was used as the second dependent measurement. In the American version of the experiment, this responsibility question was added to the survey in the middle of data collection. As a result, only 285 out of 649 American participants received the responsibility question, whereas all Korean participants received this question.

On the following page, two questions were asked as comprehension checks: “What is Sam’s relation to Brian?” and “Did Brian financially benefit from Sam’s fortune?” Next, the verbal version of a measure of participants’ identity fusion with their own family (Swann et al., 2012) was administered. This measure includes 7 items, each rated on a 7-point scale (1: strongly disagree; 7: strongly agree). Identity fusion assesses how much people align themselves with an affiliated group and experience “oneness”. In the present study, participants’ fusion with their family was measured by replacing the word “country” in Swann Jr. et al.’s questionnaire with

“family”. The degree to which people identify themselves with family was expected to predict differences in collective responsibility judgments. Finally, basic demographic information was collected.

Results

A three-way factorial ANOVA was conducted to assess the influence of culture, relationship, and financial benefit (as categorical variables) on the amount of donation to the O’Neal children’s education (see Figure 3.1). Donation amounts for Korean samples are reported after conversion to U.S. dollars. The main effects of all three independent variables were significant. The main effect of culture ($F(1, 1274) = 84.5, p < .001, \eta_p^2 = .063$) reflected greater overall mean donations for Korean ($M = \$6,996, 95\% \text{ CI} = [\$6,784, \$7208]$) as compared to American ($M = \$5,647, 95\% \text{ CI} = [\$5,436, \$5,858]$) participants. The main effect of financial benefit ($F(1, 1274) = 16.9, p < .001, \eta_p^2 = .013$) reflected greater overall mean donations when financial benefit was present ($M = \$6,629, 95\% \text{ CI} = [\$6,415, \$6,843]$) rather than absent ($M = \$5,998, 95\% \text{ CI} = [\$5,780, \$6,217]$).

The main effect of relationship ($F(3, 1274) = 9.2, p < .001, \eta_p^2 = .020$) was investigated in further detail using pairwise *t*-tests between pairs of relationship conditions for the four levels of this factor, with Holm-Bonferroni correction. Participants judged that the amount of donation to the O’Neal children’s education should be greater in the biological condition ($M = \$6,875, 95\% \text{ CI} = [\$6,596, \$7,155]$) than in the remarriage condition ($M = \$6,069, 95\% \text{ CI} = [\$5,745, \$6,393]$; $t(661) = 3.70, p = .001, d = 0.29$) and no-relation condition ($M = \$5,859, 95\% \text{ CI} = [\$5,548, \$6,170]$; $t(629) = 4.78, p < .001, d = 0.38$), but not the adoptive condition ($M = \$6,412, 95\% \text{ CI} = [\$6,106, \$6,718], p = .099$). The average recommended donation tended to be higher in the adoptive condition than in the no-relation condition, although this difference was not significant

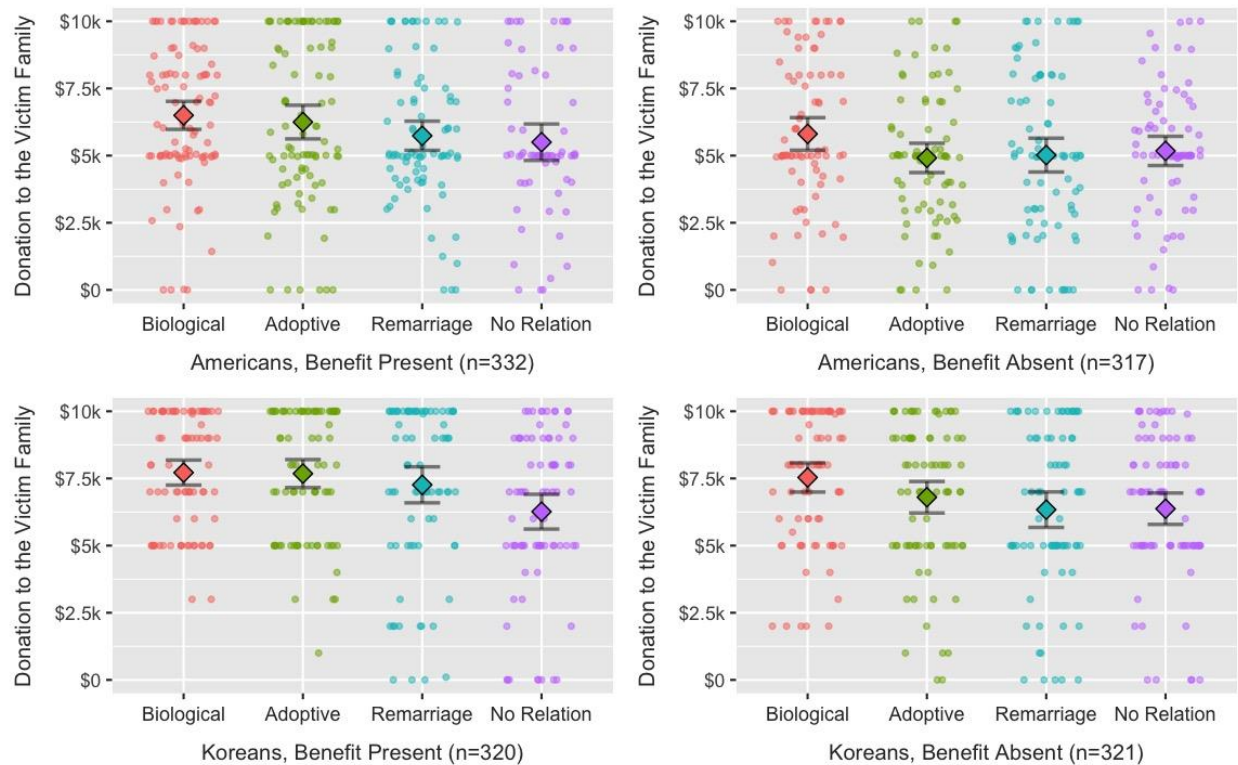


Figure 3.1. Amount of money donated to O’Neal children’s education in each of the sixteen conditions. Diamonds indicate group means. The dots show the distributions of responses across the range from \$0 to \$10,000. Error bars indicate 95% confidence interval (assuming normal distribution).

after correction ($t(625) = 2.48, p = .054$). The other three pairwise comparisons were not significant ($ps > .10$). None of the second- or third-order interaction effects were significant ($Fs < 2.06, ps > .103$). Therefore, Korean participants were not different from American participants in distinguishing the four kinds of relationships between the agent and the wrongdoer.

Next, an analogous three-way ANOVA was used to analyze predictors of the level of collective responsibility assigned to Brian (see Figure 3.2). As in the analysis of monetary donations, significant main effects of culture ($F(1, 910) = 142.8, p < 001, \eta_p^2 = .140$), relationship

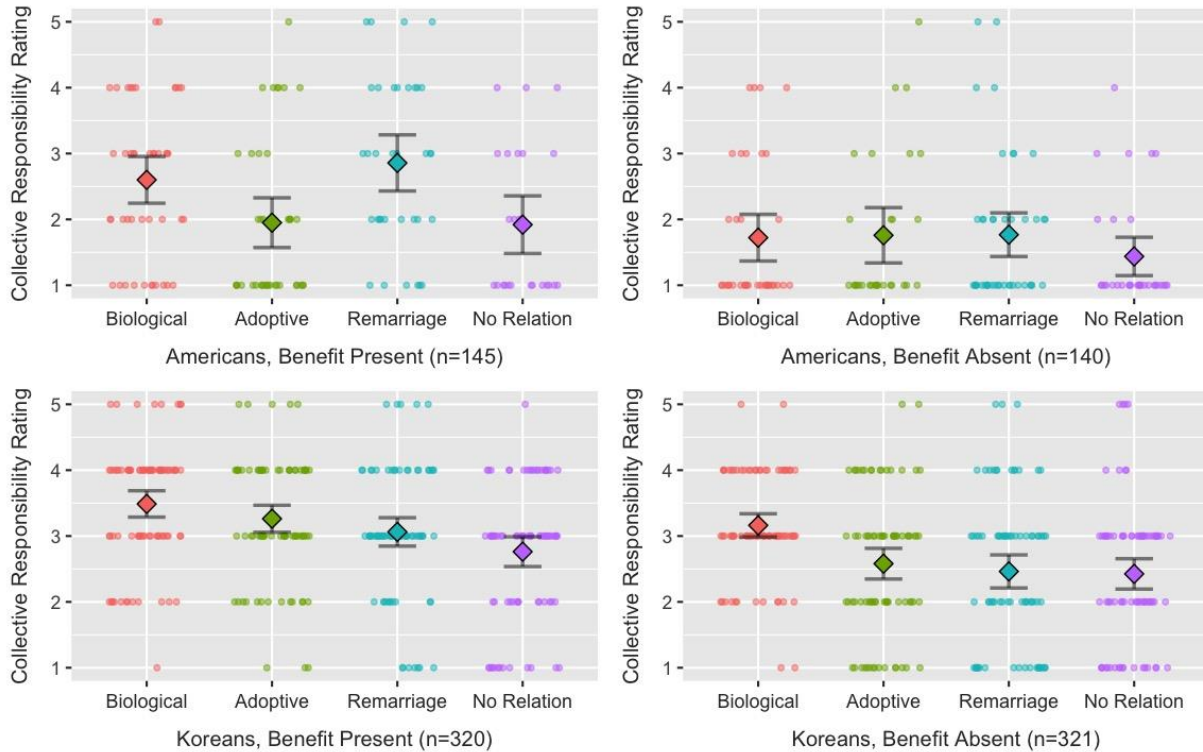


Figure 3.2. Participants' ratings that Brian should feel collective responsibility for what happened to the O'Neal family, in each of the sixteen conditions. Diamonds indicate group means. Distributions of data points across the range from 1 (*not at all responsible*) to 5 (*fully responsible*) are shown by the dots. Error bars indicate 95% confidence intervals (assuming normal distribution).

($F(3, 910) = 11.2, p < .001, \eta_p^2 = .041$), and financial benefit ($F(1, 910) = 58.1, p < .001, \eta_p^2 = .062$) were found. As predicted, Koreans ($M = 2.90, 95\% \text{ CI} = [2.82, 2.98]$) assigned generally higher responsibility to the agent than did Americans ($M = 2.03, 95\% \text{ CI} = [1.89, 2.17]$). Across both cultures, collective responsibility score was higher when financial benefit was present ($M = 2.90, 95\% \text{ CI} = [2.80, 3.01]$) than absent ($M = 2.36, 95\% \text{ CI} = [2.25, 2.47]$). In addition, the interaction

between culture and relationship was significant ($F(3, 910) = 4.8, p = .003, \eta_p^2 = .015$). No other interaction effects were significant ($F_s < 2.58, p_s > .052$).

However, the significant interaction between culture and relationship was due to unexpectedly high rating of responsibility in the remarriage condition among American participants but not among Korean participants (see Figure 3.2). Two regression analyses were conducted using U.S. \times biological \times benefit-present and U.S. \times no-relation \times benefit-present conditions as reference groups, respectively, and the benefit-absent dummy variable was entered as a first-order predictor. Results showed that among the U.S. participants, the responsibility ratings in the remarriage condition were significantly higher than in the no-relation condition ($b = 0.60, t(917) = 3.30, p = .001$), and even higher than in the biological condition ($b = 0.11, t(917) = 0.63, p = .53$). These high ratings in the remarriage condition relative to the two reference conditions significantly decreased in Korean participants, as the two-way interaction terms between dummy variables coding culture and remarriage were significant (when no-relation was the reference group, $b_{culture \times remarriage} = -0.43, t(917) = -2.0, p = .046$; when biological was the reference group, $b_{culture \times remarriage} = -0.67, t(917) = -3.28, p = .001$). Other interaction terms were not significant ($p_s > .27$). This difference in the pattern may reflect the less reliable estimation of group means in the U.S. dataset due to smaller sample size.

Pairwise t -tests between relationship conditions were performed as a follow-up to the significant main effect of relationship after collapsing the responses across cultures. Using Holm-Bonferroni correction, assigned collective responsibility was higher in the biological condition ($M = 2.95, 95\% \text{ CI} = [2.81, 3.09]$) than in the other three conditions: adoptive ($M = 2.60, 95\% \text{ CI} = [2.45, 2.76]; t(469) = 3.22, p = .005, d = .30$), remarriage ($M = 2.60, 95\% \text{ CI} = [2.44, 2.75]; t(477) = 3.32, p = .005, d = .30$), and no-relation ($M = 2.35, 95\% \text{ CI} = [2.20, 2.49]; t(456) = 5.74, p < .001$,

$d = .54$). Compared to the no-relation condition, adoptive ($t(445) = 2.36, p = .057$) and remarriage ($t(453) = 2.31, p = .057$) conditions scored higher, but the differences were not significant after correction. The difference between adoptive and remarriage conditions was not significant ($t(466) = 0.07, p = .94$).

I checked for any changes in the pattern of results after filtering out Koreans who self-reported a prolonged exposure to foreign culture (23%). The statistical pattern remained the same for both ANOVAs after removing such participants.

The difference in identity fusion with family (average of 7 ratings; $\alpha = .94$) between American ($M = 4.96, 95\% \text{ CI} = [4.85, 5.06]$) and Korean ($M = 5.28, 95\% \text{ CI} = [5.18, 5.37]$) participants was significant ($t(1,288) = 4.32, p < .001, d = 0.24$). Mediation analyses using PROCESS macro version 3.0 for SPSS (Hayes, 2018) were conducted to test if the cultural difference on the dependent measurements is mediated by identity fusion. For these analyses, no-relation conditions (137 Americans and 160 Koreans) were excluded because identity fusion with family was expected to covary with the dependent measurements only when the scenario described the wrongdoer and the agent as members of the same family. Of the 285 American participants who answered the collective responsibility question, 57 were in the no-relation condition and were excluded.

Pre-defined model 4 in PROCESS was used. Three dummy variables coding the levels of independent variables (indicating adoptive, remarriage, and benefit-present conditions, respectively) were entered as covariates for predicting the mediator (identity fusion score) and the dependent measurements. Indirect effects were estimated through bootstrapping ($n = 20,000$). The direct effect of culture on the suggested amount of donations was significant ($b = \$1,516, t(987) = 8.81, p < .001, 95\% \text{ CI} = [\$1,178, \$1,854]$), while the indirect effect through identity fusion score

was not ($b = -\$2.6$, 95% CI = $[-\$48.8, \$41.0]$). When predicting the collective responsibility rating, the direct effect of culture ($b = 0.89$, $t(703) = 10.36$, $p < .001$, 95% CI = $[0.72, 1.05]$) and the indirect effect through identity fusion score ($b = 0.02$, 95% CI = $[0.00, 0.05]$) were both significant, although the latter was small in size. Thus, there was limited evidence to support the hypothesis that the effect of culture on collective moral judgments is mediated by participants' identification with their family.

Study 3B

Study 3A showed that there is a significant difference across societies in assigning collective responsibility. In addition, the main effects of relationship between wrongdoer and agent (level of entitativity of the wrongdoer group) were significant.

As a follow-up experiment, Study 3B addressed a number additional questions: (a) whether the same pattern holds when the scenario is written in first-person perspective, (b) whether the same pattern holds for groups other than family, (c) whether the effect of culture can be replicated in Japan, another collectivistic East Asian country, in addition to the U.S. and Korea, (d) whether collective pride, as a reaction to another group member's virtuous action, exhibits a similar pattern as collective guilt and responsibility, and (e) whether participants' identity fusion with a particular group and/or their general collectivism score mediate the influence of culture on moral judgments, when the scenarios have a first-person narrative.

Method

Participants and Design

A 2 (culture: U.S./South Korea and Japan) \times 3 (actor's relation to participants: brother/friend/acquaintance) \times 2 (moral behavior: harm/help) design was used, where all independent variables were manipulated between subjects. Note that we additionally manipulated

the valence of moral behavior of the actor between subjects in this study, orthogonally to the other two independent variables. That is, half of the participants read about an actor harming a person, whereas the other half read about the same actor helping a person. Each participant read and responded to one scenario.

Because the responses from the harm and help conditions were to be analyzed separately, power analysis for a planned 2 (culture: U.S./South Korea and Japan) \times 3 (actor's relation to participants: brother/friend/acquaintance) ANOVA was conducted using R package Superpower (Lakens & Caldwell, 2021). To detect the smallest effects observed in Study 3A ($\eta_p^2 = .06$, Cohen's $d = 0.37$) with the power of .80 and α of .05, at least 28 participants were needed in each of the 6 conditions. To detect more fine-grained distinctions between the three relationship conditions with high power, it was predicted that about 40 participants will be needed in each of the six conditions. In addition, to prepare for the possibility of Japanese and South Korean participants showing significantly distinctive patterns of responses, 40 participants were recruited for each of the three relationship conditions in all three countries. As a result, more than 240 participants were recruited from the three countries.

Amazon MTurk was used to recruit 257 participants residing in the U.S. (109 females, $M_{age} = 37.5$, $SD_{age} = 8.5$); Hankook Research recruited 242 participants in Korea (124 females, $M_{age} = 36.6$, $SD_{age} = 12.3$); Lancers, Inc. (<https://www.lancers.co.jp/>) recruited 252 participants in Japan (115 females, $M_{age} = 37.5$, $SD_{age} = 11.1$). Payment rates were higher than the minimum hourly wage in each country. Average participation times were between 8 and 11 minutes. Additional 99 American and 8 Japanese participants were recruited but excluded from all analyses because they did not pass one of the two basic attention checks. American and Japanese participants who did not pass both attention checks were automatically dropped from the experiment as soon as they

failed the second check. Korean participants who failed at least one attention check were automatically dropped from the experiment in the survey agency's internal system.

Materials and Procedure

Korean and Japanese versions of the scenarios were first translated from the English version by the authors and then back translated by two independent bilingual speakers (Korean-English and Japanese-English bilinguals) per version to check for tone and content. Through discussions between the authors and the bilingual translators, any potential discrepancy between different versions of scenarios were resolved. The actor was referred to as 'B' in the friend and acquaintance conditions, and there were no proper nouns indicating names of individuals or locations that needed to be translated in Study 3B,

As in Study 3A, participants were instructed to read a hypothetical scenario and answer a few related questions. In the scenarios, an actor who exhibits either a harmful or helpful moral behavior was described as participants' brother, a close friend from school, or an acquaintance, depending on the relationship condition. Participants were instructed to think about their actual brother, friend, or acquaintance while reading the scenario, to increase their engagement with the situation. If they did not have a brother, they were asked to imagine that they have a brother who is a few years older than them.

In the scenarios for the brother and friend conditions, participants met with their family or a close group of friends (including the actor 'B') to take a short vacation together at a beachside hotel. In the acquaintance condition, the scenario supposed that participants discovered a casual meet-up group on social media that hosts movie nights for anyone willing to join. Participants realized that the online group will soon be watching a movie that they were very interested in at a hotel, and they decided to go to the event. The only member of the online group they knew was an

acquaintance ‘B’, who had been briefly introduced to participants through a mutual friend in the past.

The actor in each scenario told participants that he will join the event later because he first wanted to stop by at a friend’s place. On the first night of the vacation/event, participants received a phone call from the actor. In the harm condition, the actor confessed that he hit a pedestrian on his way to the hotel, and he had drunk wine before driving. Participants calmed him down, and soon he was arrested by the police. It was revealed that the victim was a young woman and she fell in a coma with no sign of recovery due to the accident, leaving her family in great distress. In the help condition, the actor said that he found a middle-aged woman having a seizure on the sidewalk and quickly drove her to a nearby hospital, saving her life. A few days later, the woman’s family expressed gratitude to the actor and the participants’ group in tears.

In all scenarios, the actor and participants were members of a group that had a common goal and intention of taking a vacation or watching a movie. This premise was intended to elicit a sense of entitativity (Campbell, 1958; Hamilton et al., 2002) even in the acquaintance condition where participants and members of the online group did not previously share categorical group membership (Rutchick et al., 2008). At the same time, the moral behavior of the actor was committed without the participants making any causal contribution to it. Therefore, analogous to Study 3A, Study 3B created situations where participants (analog of Brian in Study 3A) could experience collective moral emotions or take vicarious responsibility for the behavior of another member of an in-group.

After reading the scenario, participants in the harm condition rated the degree to which they agree with the statements describing collective guilt (“Even though I was not involved, I feel guilty for what [actor] did to the woman and her family”, “As a member of [group], I feel guilty

when I think about what [actor] did to the woman and her family”, $r = 0.77$), feeling of responsibility (“Even though I was not involved, I feel responsible for what [actor] did to the woman and her family”), and the urge to apologize (“I feel that I should apologize for what [actor] did”). In the help condition, participants were shown statements describing collective pride (‘guilty’ in the two collective guilt questions were replaced with ‘proud’, $r = 0.40$) and feeling of contribution (“Even though I was not involved, I feel like I contributed to the behavior of [actor]”). The responses to these five questions were used as main dependent measurements in this study. All questions were answered on a 5-point Likert scale (1: *completely disagree*, 5: *completely agree*). An additional question was asked in both the harm and help conditions (feeling of involvement: “I feel like [actor] and I were involved in the described incident together”). However, Japanese participants showed distinct patterns of responses from American and Korean participants in these questions, which could not be easily explained by the variables measured in this study. Therefore, the current section will focus on the results involving the five dependent measurements described earlier.

Two manipulation check questions were asked on the following page to ensure that participants were aware of the two critical manipulations about the relationship and the valence of behavior shown by the actor (“What was the relationship between you and the man appearing in the scenario? – (a) He was your brother, (b) He was your close friend), (c) He was a member of an online group you recently joined”, “What did the man in the scenario do? – (a) He harmed a person, (b) He helped a person”).

Then, several tests of individual differences were administered, including identification with the group that appeared in the scenario (Tropp & Wright, 2001), group-specific collectivism scores (Chen & West, 2008), and the general collective identity orientation score (Cheek et al.,

2002). For Tropp and Wright's (2001) test, participants chose one out of seven diagrams that best represented the level of overlap in identities between themselves and the group that appeared in the scenario. Study 3A used verbal questionnaire for identity fusion with group (Swann et al., 2012), which is conceptually similar to Tropp and Wright's (2001) test. Because the scores from verbal questions did not strongly correlate with the dependent variables, a shorter, visual version of the test was administered. For Chen and West's (2008) questionnaire, participants reported their tendency to share positive and negative outcomes with the relevant in-group in their daily lives (e.g., "If I were successful, [group] members would be honored") and how much they consider the implications of their decisions for the in-group (e.g., "When making decisions, it is important for me to take [group] members' feelings into account"). The questionnaire from Cheek et al. (2002) did not measure participants' attitudes toward the groups that appeared in the scenario, but asked how important various social groups are (e.g., Being a part of many generations of my family, ethnic background, being a citizen of one's country) to the sense of who they are. These scores were collected for exploratory analyses to discover potential mediating factors between culture and collective moral judgments.

Results

Overall, the dependent measurements showed similar distributions in Korean and Japanese participants. Therefore, a two-way factorial ANOVA was used to analyze each of the four dependent variables as a function of culture (Americans vs. East Asians) and relationship (see Figures 3.3 and 3.4 for trends across the three countries). Because the pattern of responses observed in the help conditions was straightforward and less contentious, participants' judgments of collective pride and feeling of contribution in the help conditions will be discussed first. Then, the

Figure 3.3. Participants' experience of collective pride (top row) and feeling of contribution to the actor's moral behavior (bottom row) as functions of relationship type and country. Diamonds and error bars indicate group means and 95% confidence intervals, respectively.

responses to the collective guilt and responsibility questions in the harm conditions will be discussed in comparison to the findings from other studies and the help conditions.

In the help conditions (see Figure 3.3), the type of relationship significantly predicted both collective pride and feeling of contribution ($F_s > 3.13$, $p_s < .045$, $\eta_p^2 = .14$ for collective pride; $\eta_p^2 = .02$ for feeling of responsibility). Collapsing across countries, reported collective pride was from high to low in the order of brother ($M = 4.51$, 95% CI = [4.41, 4.61]), friend ($M = 4.30$, 95% CI = [4.19, 4.41]), and acquaintance ($M = 3.84$, 95% CI = [3.71, 3.95]), as expected (pairwise t -test $p_s < .009$ with Holm-Bonferroni correction). As for feeling of contribution, the same order was not

Figure 3.4. Participants' experience of collective guilt (top row), feeling of responsibility for the actor's immoral behavior (middle row), and the urge to apologize (bottom row), as functions of relationship type and country. Diamonds and error bars indicate group means and 95% confidence intervals, respectively.

clearly observed in the pairwise comparisons between brother ($M = 3.34$, 95% CI = [3.09, 3.59]), friend ($M = 3.05$, 95% CI = [2.79, 3.31]), and acquaintance ($M = 3.14$, 95% CI = [2.90, 3.38])

conditions ($ps > .31$ with correction). Meanwhile, the main effect of culture ($F_s < 3.11, ps > .078$) and the interaction between relationship and culture ($F_s < 2.10, ps > .12$) were not significant for either of the two dependent measurements. In sum, when given a chance to experience positive moral emotions such as pride and feeling of contribution, cross-cultural differences did not emerge. Because they could endorse a positive self-view (Alicke & Sedikides, 2009) without any cost, risk, or social pressure to do otherwise, participants seem to have chosen to do so to a similar degree across countries.

In the harm conditions (see Figure 3.4), collective guilt (average of two scores), feeling of responsibility, and the urge to apologize were all significantly influenced by the type of relationship ($F_s > 46.18, ps < .001, \eta_p^2s > .17$) and the interaction between relationship and culture ($F_s > 6.67, ps < .001, \eta_p^2s > .03$). However, the main effect of culture was not significant ($F_s < 1.66, ps > .19$). As in the help conditions, the effect of type of relationship was observed in the expected direction regardless of culture, such that participants experienced higher level of collective guilt and responsibility when they associated themselves with the actor more strongly (brother > friend > acquaintance after collapsing scores across countries, $ps < .035$ with Holm-Bonferroni correction). However, contrary to the prediction based on cross-cultural studies (Bedford & Hwang, 2003; Oyserman et al., 2002; Triandis, 2001; Triandis & Gelfand, 1998) and the results from Study 3A, East Asian participants reported generally similar or lower levels of collective guilt ($M = 3.23, 95\% \text{ CI} = [3.09, 3.38]$), feeling of responsibility ($M = 2.98, 95\% \text{ CI} = [2.83, 3.13]$), and urge to apologize ($M = 2.59, 95\% \text{ CI} = [2.43, 2.75]$) than American participants (collective guilt: $M = 3.47, 95\% \text{ CI} = [3.26, 3.68]$; feeling of responsibility: $M = 3.06, 95\% \text{ CI} = [2.82, 3.31]$; urge to apologize: $M = 3.33, 95\% \text{ CI} = [3.10, 3.56]$). Meanwhile, the significant two-way interaction suggested that, compared to American participants, East Asian participants may

be making clearer distinctions in judgments based on the type of relationships between themselves and the actor: In their reports of collective guilt and feeling of responsibility, East Asians' average scores in the brother ($ts < 1.40, ps > .16$) and acquaintance ($ts < 1.72, ps > .089$) conditions were similar with Americans, but their scores in the friend ($ts > 2.36, ps < .020$) condition were significantly lower than Americans; In their reports of the urge to apologize, Americans and East Asians scored similar in the brother condition ($t(125) = 0.73, p = .47$), but East Asians scored much lower than Americans in the friend ($t(122) = 2.85, p = .005, d = 1.18$) and acquaintance ($t(124) = 3.45, p < .001, d = 0.62$) conditions.

It was surprising that East Asian participants reported lower scores than Americans in the harm conditions. A simple interpretation of collectivism (Triandis, 2001; Triandis & Gelfand, 1998) suggests that moral judgments made by members of a collectivistic society may be more influenced by group affiliations of individuals involved in a situation and relevant intergroup dynamics.

However, a closer look at the individual difference measurements provided insights into the unexpected cultural differences. All three measurements (identification with group, group-specific collectivism scores, and general collective identity orientation) showed similar patterns of difference between American and East Asian participants. For simplicity, the current section will only use group-specific collectivism scores (18 items, as of each of the three subcomponents > 0.8 , Chen & West, 2008; see Figure 3.5). In a 2 (culture: Americans vs. East Asians) \times 3 (actor's relation to participants: brother/friend/acquaintance) \times 2 (valence of actor's moral behavior: harm/help) ANOVA to predict the group-specific collectivism scores, the two-way interaction between culture and relationship was significant ($F(2, 739) = 6.05, p = .002, \eta_p^2 = .02$), while other independent variables that included the culture variable were not ($F_s < 2.34, ps > .097$).

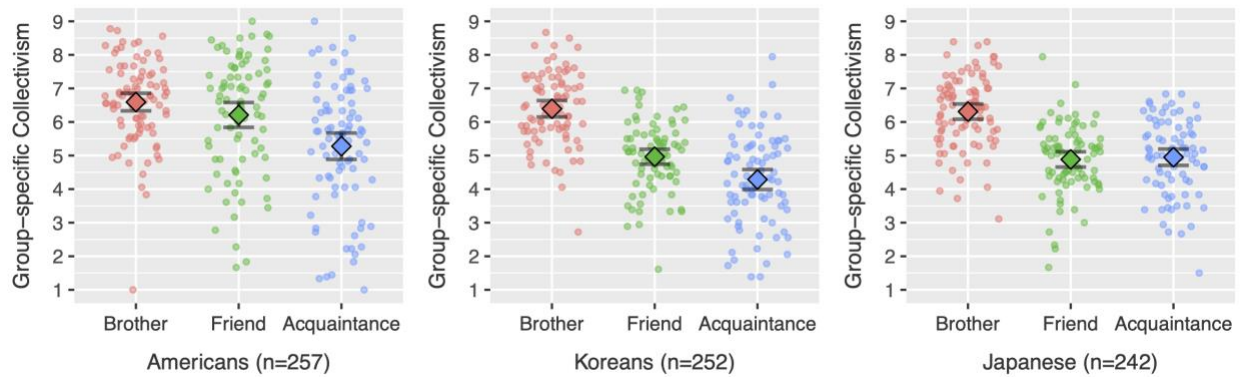


Figure 3.5. Group-specific collectivism scores (Chen & West, 2008) reported by participants from the U.S., South Korea, and Japan. Diamonds and error bars indicate group means and 95% confidence intervals, respectively.

Follow-up *t*-test revealed that Americans and East Asians did not significantly differ in their willingness to share life outcomes with family or consider the impact of their decisions on family members ($t(255) = 1.61, p = .11$). However, Americans scored significantly higher than East Asians when reporting their willingness to do the same for the group of friends ($t(243) = 7.29, p < .001, d = 0.91$) or the online group ($t(247) = 3.29, p = .001, d = 0.41$) that appeared in the scenario.

Mediation analyses were conducted in PROCESS using the same configuration as the analyses from Study 3A, to test whether the effect of culture (Americans vs. East Asian) on each of the dependent measurements was mediated by the group-specific collectivism scores (see Figure 3.6 for path models of four of the five dependent measurements). ‘Culture’ variable in the mediation analyses was a dummy variable that coded American participants as 0 and East Asians as 1. Therefore, the effect of ‘culture’ can be interpreted as the difference between Americans and East Asians where a negative sign indicates lower scores reported by East Asians. Moreover, acquaintance condition was used as a reference group and two dummy variables coding the other

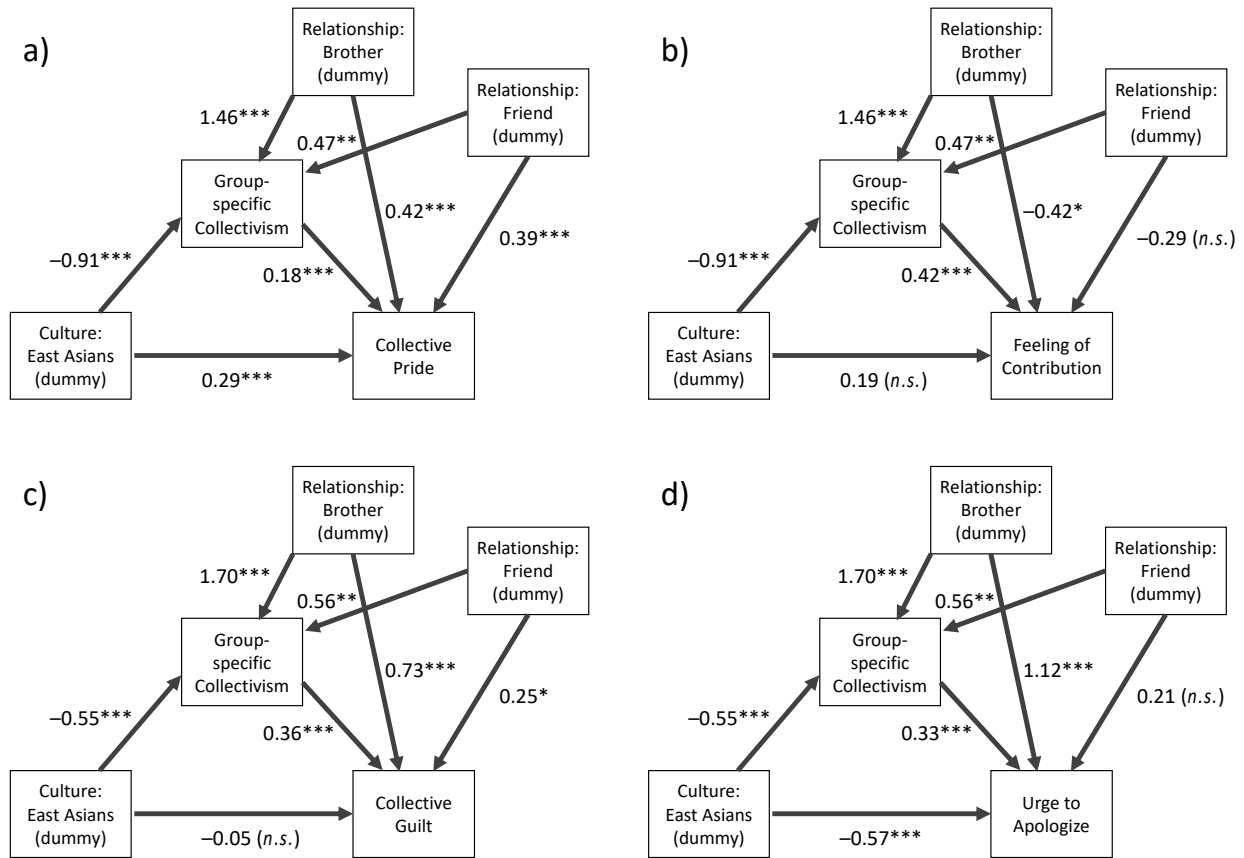


Figure 3.6. Mediation analyses to test if the effect of culture (Americans vs. East Asians) on each of the dependent measurements is mediated by group-specific collectivism scores, while controlling for the effect of relationship manipulations. The four graphs show analyses for (a) collective pride, (b) feeling of contribution, (c) collective guilt, and (d) urge to apologize. Weights indicate unstandardized regression slopes, and the negative weights for arrows stemming from ‘Culture: East Asians (dummy)’ box indicate lower scores reported by East Asians compared to Americans. * $p < .05$, ** $p < .01$, *** $p < .001$.

two relationship types (brother and friend conditions) were used as covariates that predict the mediator (collectivism scores) and the dependent variables.

In the help condition, the indirect effect was estimated to be significantly lower than zero for both collective pride ($b = -0.16$, 95% CI = $[-0.23, -0.10]$) and feeling of contribution ($b = -0.39$, 95% CI = $[-0.55, -0.24]$). The direct effect of culture after ruling out the indirect effect was still significant for collective pride ($b = 0.29$, $t(369) = 4.29$, 95% CI = $[0.15, 0.42]$, $p < .001$) but not for feeling of contribution ($b = 0.19$, $t(369) = 1.23$, 95% CI = $[-0.11, 0.49]$, $p = .22$).

In the harm condition, the indirect effect was estimated to be significantly lower than zero for collective guilt ($b = -0.20$, 95% CI = $[-0.33, -0.07]$), feeling of responsibility ($b = -0.21$, 95% CI = $[-0.36, -0.06]$), and urge to apologize ($b = -0.18$, 95% CI = $[-0.31, -0.06]$). The direct effect of culture distinguished from the mediated effect was still significant for urge to apologize ($b = -0.57$, $t(372) = 5.26$, 95% CI = $[-0.79, -0.36]$, $p < .001$), but not for collective guilt ($b = -0.05$, $t(372) = 0.55$, 95% CI = $[-0.25, 0.58]$, $p = .58$) or feeling of responsibility ($b = 0.12$, $t(372) = 1.06$, 95% CI = $[-0.10, 0.29]$, $p = .29$).

General Discussion

Assignment of collective responsibility based on group membership deviates from the traditional moral norms supported as philosophical principles within the individualistic Western societies (Smiley, 2017; Talbert, 2019). Nevertheless, laypeople across societies seem to take group membership into account when making moral judgments. Moreover, they seemed to believe that a varying degree of collective guilt, pride, and responsibility can be experienced by themselves and assigned to others depending on the type of group involved. Study 3A found that for both American and Korean participants, normative judgments of collective responsibility were greater when an agent in a hypothetical scenario had received financial benefit traceable to the wrongdoer. Participants also demonstrated sensitivity to the type of relation between the wrongdoer and the

agent. A similar effect of relationship type was observed in Study 3B where Americans, South Koreans, and Japanese participants made moral judgments from a first-person perspective.

The results demonstrate an intriguing exception to the connection between causality and judgment of responsibility. Psychological theories of morality have postulated that a causal relation between an agent's behavior and harm is critical for an observer to assign blame and moral responsibility to the agent (e.g., Cushman, 2008; Waldmann & Dieterich, 2007). The influence of non-causal factors (group identity, moral obligation, social norms; Ditto et al., 2009; Holyoak & Powell, 2016; Malle et al., 2014) remains to be incorporated into rational computational models of moral judgments.

However, based on the previously validated scales of collectivism, the East Asian participants recruited in Study 3B did not appear to be more collectivistic than the American participants. In Study 3A, it seemed that Korean participants were showing a slightly higher level of identity fusion with their family (Swann et al., 2012) than American participants. However, an extended design in Study 3B showed that when comparing people's level of involvement with or concern for different social groups, Americans appeared more collectivistic than East Asians in thinking about those groups. Hence, the general prediction that individuals with stronger collectivistic values will report stronger experience of collectivistic guilt, pride, and responsibility was supported by data—except in the opposite direction.

Why were the East Asians in Study 3B less collectivistic than Americans, contrary to the past findings? Clearly, increasing emphasis on individualistic practices and values has been an ongoing trend across the world since the mid-20th century, which often followed socioeconomic development of traditionally collectivistic countries (Hamamura, 2012; Santos et al., 2017). In this light, it is possible that since the foundational studies in cross-cultural psychology first examined

the East-West differences (e.g., Nisbett et al., 2001; Triandis, 2001; Triandis & Gelfand, 1998), the gap between East Asia and the Western societies in terms of individualism-collectivism has been closing.

But recall that Koreans in Study 3A imposed much greater collective responsibility to make reparations to another agent than Americans did. The key distinction between the two studies was whether participants made moral judgments from a third-person or first-person perspective. It seems that East Asians were more inclined to impose collective responsibility to others compared to Americans, but they were hypocritically less inclined than Americans to take collective guilt and responsibility themselves.

A systematic explanation for these diverging results may be proposed based on the substantive difference in social norms accepted in different cultures. For example, research on the cross-cultural difference along the tightness-looseness dimension (Eriksson et al., 2021; Gelfand et al., 2011) showed how tight cultures with strong social norms have stricter regulations and prescriptions for behaviors of individuals and are less permissive of deviant behaviors compared to loose cultures with weaker norms. Although our current studies suggest that East Asians in today's world may not be more collectivistic than Americans, East Asian countries still seem to have much tighter cultures than the U.S. or other Western countries (Gelfand et al., 2011).

According to Bicchieri (2005), norms are not necessarily what are overtly and frequently observed in people's behaviors. Rather, they are our understanding of what behaviors and beliefs are accepted and expected by other members of the society. Individuals comply with social norms most of the times, but they sometimes exploit their understanding of these expectations to pretend that they are following the norms even though they are in fact violating them to maximize personal utility. A similar process may be at work when members of a tight culture assign collective

responsibility to themselves and others. For example, in order to receive the benefits of a society with strong social norms (e.g., better coordination and orderliness when facing crises such as the Coronavirus pandemic, Gelfand et al., 2021), individuals must participate in the collective effort to strictly regulate the behaviors of members of the society and maintain order. However, in situations where their own behaviors can potentially be regulated or blamed by others, members of a tight culture may desire to avoid them as much as possible, because the ensuing liability and possible punishment is particularly strong and demanding in their society (Gelfand et al., 2011).

Therefore, future studies may test the idea that first-person collective moral judgments are shaped by people's personal identification with groups (as in Study 3B) whereas third-person collective moral judgments are shaped by how much they endorse strong social norms that impose obligations and responsibilities to others (which may have been hinted in Study 3A). Identification with group or group-specific collectivism scores were shown to mediate the effect of culture on moral judgments only when judgments were made in first-person perspective and not when they were made in third-person perspective. These data provide evidence that there is an important difference between the psychological processes of making moral judgments about oneself versus others (e.g., Hirschfeld-Kroen et al., 2021), which have been largely understudied in moral psychology.

It should be stressed that cross-cultural psychologists have long warned about the danger of dichotomizing the societies into categories labeled as collectivistic and individualistic (Oyserman et al., 2002; Vignoles et al., 2016). Even if there exists a significant influence of culture on psychological processes, its impact is likely to interact with types of social situations and relationships, which obscures the interpretation of such effect (Chen & West, 2008; Oyserman et al., 2002). To overcome such difficulty in cultural psychology, future theories will need to consider

not only laypeople's understanding of individual and group identity but also the social norms that fill in our social life with detailed schemas for thinking about and interacting with others.

Conclusion

Our inclination to achieve coherence in reasoning and perception of the world plays a critical role in shaping our thoughts and behaviors. On one hand, maintaining logically consistent, parsimonious, and appealing sets of explanations to understand the observed events in our lives greatly helps us make predictions and control the environment (Cheng & Lu, 2017; Lombrozo, 2016; Powell et al., 2016). On the other hand, the desire to construe the ambiguous world in a coherent and efficient manner sometimes produces biases that may not be rational or moral (Kunda, 1990; Kunda & Thagard, 1996; Holyoak & Simon, 1999; Simon et al., 2015; Roediger & McDermott, 1995).

In the introduction, I labeled one interpretation of coherence that focuses on logical consistency and explanatory relations as rational coherence (Cheng & Lu, 2017; Marr, 1982; Powell et al., 2016; Thagard, 1989) and another interpretation that focuses on attitudes, motivations, emotions, and identities as cognitive consistency (Festinger, 1957; Heider, 1946; Kunda, 1990). Both interpretations were useful in understanding how people reason about complicated issues that require considerations of multiple factors.

Cognitive consistency theories provided the framework for representing an intricate system of motivations, emotions, beliefs, and self-image. Laypeople's reasoning process is clearly influenced by emotional factors even when rational analysis does not warrant it. Therefore, an accurate descriptive model of reasoning should be capable of encoding various psychological

constructs ranging from the components of ‘hot cognition’ (e.g., emotion, motivation, preference; Simon et al., 2015; Thagard, 2008) to scientific hypotheses (e.g., Thagard, 1989) to moral rules (e.g., rights and duties; Holyoak & Powell, 2016). Cognitive consistency theories also provided predictions for situations where conflicting beliefs, judgments, and self-identities coexist. A dynamic shift to resolve the imbalance (Heider, 1946; Holyoak & Simon, 1999; Simon et al., 2015) occurred in some of the reported studies.

The theories of rational coherence were particularly useful in designing interventions and educational materials that aim to change beliefs and attitudes. The findings from causal learning literature provided a systematic account of how laypeople’s beliefs and causal understanding of the world may be revised without backfiring (Cheng, 1997; Cheng & Lu, 2017). Moreover, the rational view of coherence stresses that reasoners—including people with extreme or unpopular views such as conspiracy theorists—strive to organize their beliefs and knowledge in a parsimonious way, by creating narratives and adopting worldviews (Heine et al., 2006; Hochschild, 2018; Ochs & Capps, 2001). By acknowledging this rational aspect of what people commonly dismiss as irrational thinking by out-group members, our society may be able to foster a civil discussion about many contentious issues.

Using coherence in reasoning as the overarching theme, the current set of studies attempted to understand how people make judgments and decisions in relation to important social issues in today’s world. These studies showed different ways in which nuanced real-world factors (e.g., cultural norms, conservative identities) and phenomena (historical injustice, #MeToo, climate change denialism) can be combined with traditional cognitive theories to explain our reasoning about complicated issues. Extensions of these ideas to related issues and topics will open the door to exciting new ways of understanding social phenomena and production of pragmatic

interventions that can accelerate positive social change. However, more theoretical work is needed to clarify the mechanisms underlying the formation, maintenance, and revision of beliefs.

Finally, I will suggest some ideas for future work as follow-ups to the studies reported in this dissertation. An obvious extension of Study 1 is to apply the same principles of cognition to produce interventions for belief revision in other domains. For example, the severe polarization of discussions surrounding gun control, women's reproductive rights, discrimination against minority groups, and the anti-vaccination movement calls for effective education and persuasion strategies to facilitate social progress. As illustrated in Study 1, a combination of carefully designed materials can lead people to choose a more coherent and parsimonious worldview over their previous intuitive worldview, without threatening their moral identity. For instance, an intervention comprised of analogous prediction exercises could be developed that proposes a coherent, moral narrative to persuade American conservatives to be more supportive of gun-control measures, such as universal background checks. It will be crucial to provide statistical evidence to counter misconceptions about the gun culture and policies (e.g., belief that if more people carry guns, their community will be safer), while also assuring them that, owing to the safer living conditions and matured social infrastructure in today's society, the probability that they will need a gun to protect themselves is extremely low. The willingness to own firearms is likely driven by the perception of the world as a dangerous and uncertain place (Buttrick, 2020; Heine et al., 2006). Therefore, convincing gun owners and conservatives to reevaluate the utility of owning guns, and guiding them to find alternative ways to feel secure in their everyday lives, may encourage revision of their entrenched beliefs about gun control.

Because the materials used in Study 1 were deliberately designed to avoid eliciting conservative participants' aversion toward climate change and scientists, it was difficult to assess

what kind of change occurred in their worldview or beliefs. For instance, it was assumed that the effect of the full-coherence intervention critically relies on whether participants' sense of positive moral self is activated in the moral identity module. However, Study 1 did not directly measure the activated self-image or the narrative participants used to understand events in the world. Future studies may attempt to measure these variables to provide clearer evidence that our intervention operates mainly at the broader level of narratives and worldviews, rather than at the lower level of individual beliefs and emotions. For instance, the degree to which a participant views themselves as a morally responsible and caring person could be measured after the intervention in all conditions. In addition, participants may be asked to write a short essay that describes their views on the conflict between environmental conservation and support for free market/economic development (Cook & Lewandowsky, 2016). Participants' responses can be evaluated by blind coders to test whether the full-coherence intervention activates a pro-science, pro-environmental narrative.

In Study 2, the interplay between two influential dimensions of social judgment—competence and morality—was examined. The classic halo effect (Nisbett & Wilson, 1977) does not yield predictions about the varying influences of different variables on impression formation. However, recent studies have argued that the moral character of a person can have a particularly strong influence on other dimensions of person evaluation, compared to other notable variables such as competence and sociability (Brambilla et al., 2019; Goodwin et al., 2014; Uhlmann et al., 2015). A formal investigation of the bidirectional influences between variables may be possible using standardized behavior descriptions of individuals. For instance, how large a coherence shift occurs across various dimensions of social judgment following a one-unit increase (e.g., on a 7-point Likert scale) in morality/competence/likability ratings may be measured and compared. This paradigm may provide a more comprehensive picture of person evaluation and impression

formation, which are some of the typical domains in which coherence shifts are observed. Asymmetries in the bidirectional influences of variables in social judgment may be modeled using a constraint satisfaction network (Kunda & Thagard, 1996; Simon et al., 2015) to observe the different roles played by different variables in the network (e.g., does moral character judgment play a more central role in the network than other variables, such as competence ratings?)

As discussed earlier, the findings from Studies 3A and 3B can be extended to test the influence of adherence to tight vs. loose social norms (Gelfand et al., 2011) on responsibility judgment from a third-person perspective. A follow-up study may maintain the experimental design from Study 3B while in addition manipulating whether participants make moral judgments from a first-person or third-person perspective. The tightness-looseness score is hypothesized to account for cross-cultural differences in third-person moral judgments, while participants' identification with group is expected to account for cross-cultural differences in first-person moral judgments. This contrast will elucidate a more detailed picture of the dynamic interactions between abstract values, motivations, identities, and moral judgments.

References

- Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*(5), 751–763.
<https://doi.org/10.1037/0022-3514.93.5.751>
- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*(4), 556–574. <https://doi.org/10.1037/0033-2909.126.4.556>
- Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*(1), 1–48.
<https://doi.org/10.1080/10463280802613866>
- Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41*(3), 258–290. <https://doi.org/10.1037/h0055756>
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences, 115*(37), 9216–9221. <https://doi.org/10.1073/pnas.1804840115>
- Bain, P. G., Hornsey, M. J., Bongiorno, R., & Jeffries, C. (2012). Promoting pro-environmental action in climate change deniers. *Nature Climate Change, 2*, 600–603.
<https://doi.org/10.1038/nclimate1532>
- Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, A. P. (2015). Moral Judgment and Decision Making. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making* (pp. 478–515). John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9781118468333.ch17>

- Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imagining how you would feel. *Personality and Social Psychology Bulletin*, 23(7), 751–758. <https://doi.org/10.1177/0146167297237008>
- Bedford, O., & Hwang, K.-K. (2003). Guilt and Shame in Chinese Culture: A Cross-cultural Framework from the Perspective of Morality and Identity. *Journal for the Theory of Social Behaviour*, 33(2), 127–144. <https://doi.org/10.1111/1468-5914.00210>
- Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.
- Bocarnea, M. C., & Brown, W. J. (2007). Celebrity-Persona Parasocial Interaction Scale. In R. A. Reynolds, R. Woods, & J. D. Baker (Eds.), *Handbook of Research on Electronic Surveys and Measurements* (pp. 309–312). Idea Group Reference. <https://doi.org/10.4018/978-1-59140-792-8.ch039>
- Brambilla, M., Carraro, L., Castelli, L., & Sacchi, S. (2019). Changing impressions: Moral character dominates impression updating. *Journal of Experimental Social Psychology*, 82, 64–73. <https://doi.org/10.1016/j.jesp.2019.01.003>
- Buttrick, N. (2020). Protective gun ownership as a coping mechanism. *Perspectives on Psychological Science*, 15(4), 835–855. <https://doi.org/10.1177/1745691619898847>
- Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science*, 3, 14–25.
- Cheek, J. M., Smith, S. M., & Tropp, L. R. (2002, February). Relational identity orientation: A fourth scale for the AIQ. *Meeting of the Society for Personality and Social Psychology*.
- Chen, F. F., & West, S. G. (2008). Measuring individualism and collectivism: The importance of considering differential components, reference groups, and measurement invariance.

- Journal of Research in Personality*, 42(2), 259–294.
<https://doi.org/10.1016/j.jrp.2007.05.006>
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104(2), 367. <https://doi.org/10.1037/0033-295X.104.2.367>
- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17(4), 391–416. [https://doi.org/10.1016/0010-0285\(85\)90014-3](https://doi.org/10.1016/0010-0285(85)90014-3)
- Cheng, P. W., & Lu, H. (2017). Causal invariance as an essential constraint for creating a causal representation of the world: Generalizing the invariance of causal power. In M. R. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning* (pp. 65–84). Oxford University Press.
- Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439–477. [https://doi.org/10.1016/0364-0213\(94\)90016-7](https://doi.org/10.1016/0364-0213(94)90016-7)
- Chudy, J., Piston, S., & Shipper, J. (2019). Guilt by association: White collective guilt in American politics. *The Journal of Politics*, 81(3), 968–981.
<https://www.journals.uchicago.edu/doi/10.1086/703207>
- Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using Bayesian networks. *Topics in Cognitive Science*, 8(1), 160–179.
<https://doi.org/10.1111/tops.12186>
- Crawley, S., Coffe, H., & Chapman, R. (2020). Public opinion on climate change: Belief and concern, issue salience and support for government action. *The British Journal of Politics and International Relations*, 22, 102–121. <https://doi.org/10.1177/1369148119888827>

- Curtin, C. M., Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M. T., Fitzpatrick, S., Gurven, M., Kanovsky, M., Kushnick, G., Laurence, S., Pisor, A., Scelza, B., Stich, S., von Rueden, C., & Henrich, J. (2020). Kinship intensity and the use of mental states in moral judgment across societies. *Evolution and Human Behavior*, *41*(5), 415–429.
<https://doi.org/10.1016/j.evolhumbehav.2020.07.002>
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380.
<https://doi.org/10.1016/j.cognition.2008.03.006>
- Cushman, F., & Young, L. (2009). The Psychology of Dilemmas and the Philosophy of Morality. *Ethical Theory and Moral Practice*, *12*(1), 9–24.
<https://doi.org/10.1007/s10677-008-9145-3>
- Davidson, D., & LePore, E. (1986). A coherence theory of truth and knowledge. *Epistemology: An Anthology*, 124–133.
- Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. *Psychology of Learning and Motivation*, *50*, 307–338. [https://doi.org/10.1016/S0079-7421\(08\)00410-6](https://doi.org/10.1016/S0079-7421(08)00410-6)
- Drummond, C., & Fischhoff, B. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 9587–9592.
<https://doi.org/10.1073/pnas.1704882114>
- Eriksson, K., Strimling, P., Gelfand, M., Wu, J., Abernathy, J., Akotia, C. S., Aldashev, A., Andersson, P. A., Andrighetto, G., Anum, A., & others. (2021). Perceptions of the

- appropriate response to norm violation in 57 societies. *Nature Communications*, 12(1), 1–11. <https://www.nature.com/articles/s41467-021-21602-9>
- Ferguson, M. A., & Branscombe, N. R. (2014). The social psychology of collective guilt. In C. von Scheve & M. Salmela (Eds.), *Collective Emotions* (pp. 251–265). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199659180.003.0017>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83. <https://doi.org/10.1016/j.tics.2006.11.005>
- Francescani, C., & Fisher, L. (2019). Bill Cosby: A timeline of his fall from “America’s Dad” to a “sexually violent predator.” *ABC News*. <https://abcnews.go.com/Entertainment/bill-cosby-trial-complete-timeline-happened-2004/story?id=47799458>
- Fraser, B. J. (1981). *TOSRA: Test of Science-Related Attitudes: Handbook*. Australian Council for Educational Research.
- Funk, C., Hefferon, M., Kennedy, B., & Johnson, C. (2019). *Trust and mistrust in Americans’ views of scientific experts*. Pew Research Center. <https://www.pewresearch.org/science/2019/08/02/trust-and-mistrust-in-americans-views-of-scientific-experts/>.
- Gelfand, M. J., Jackson, J. C., Pan, X., Nau, D., Pieper, D., Denison, E., Dagher, M., Van Lange, P. A., Chiu, C.-Y., & Wang, M. (2021). The relationship between cultural tightness–looseness and COVID-19 cases and deaths: A global analysis. *The Lancet Planetary Health*, 5(3), e135–e144. [https://doi.org/10.1016/S2542-5196\(20\)30301-6](https://doi.org/10.1016/S2542-5196(20)30301-6)

- Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliah, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Ferrer, M., Fischlmayr, I. C., ... Yamaguchi, S. (2011). Differences Between Tight and Loose Cultures: A 33-Nation Study. *Science*, *332*(6033), 1100–1104. <https://doi.org/10.1126/science.1197754>
- Giffin, C., & Lombrozo, T. (2018). An actor's knowledge and intent are more important in evaluating moral transgressions than conventional transgressions. *Cognitive Science*, *42*, 105–133. <https://doi.org/10.1111/cogs.12504>
- Goodwin, G. P. (2015). Moral character in person perception. *Current Directions in Psychological Science*, *24*(1), 38–44. <https://doi.org/10.1177/0963721414550709>
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–168. <https://doi.org/10.1037/a0034726>
- Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, *8*, 125–130. <https://doi.org/10.1016/j.copsyc.2015.09.007>
- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364–371. <https://doi.org/10.1016/j.cognition.2009.02.001>
- Guglielmo, S. (2015). Moral judgment as information processing: An integrative review. *Frontiers in Psychology*, *6*. <https://doi.org/10.3389/fpsyg.2015.01637>

- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.
- Hamamura, T. (2012). Are Cultures Becoming Individualistic? A Cross-Temporal Comparison of Individualism–Collectivism in the United States and Japan. *Personality and Social Psychology Review*, *16*(1), 3–24. <https://doi.org/10.1177/1088868311411587>
- Hamilton, D. L., Sherman, S. J., & Castelli, L. (2002). A group by any other name—The role of entitativity in group perception. *European Review of Social Psychology*, *12*(1), 139–166. <https://doi.org/10.1080/14792772143000049>
- Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication Monographs*, *85*(1), 4–40. <https://doi.org/10.1080/03637751.2017.1352100>
- Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology*, *21*(1), 107–112. <https://doi.org/10.1080/00223980.1946.9917275>
- Heine, S. J., Proulx, T., & Vohs, K. D. (2006). The meaning maintenance model: On the coherence of social motivations. *Personality and Social Psychology Review*, *10*(2), 88–110. https://doi.org/10.1207/s15327957pspr1002_1
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29. <https://doi.org/10.1038/466029a>
- Hirschfeld-Kroen, J., Jiang, K., Wasserman, E., Anzellotti, S., & Young, L. (2021). When my wrongs are worse than yours: Behavioral and neural asymmetries in first-person and third-person perspectives of accidental harms. *Journal of Experimental Social Psychology*, *94*, 104102. <https://doi.org/10.1016/j.jesp.2021.104102>

- Hochschild, A. R. (2018). *Strangers in their own land: Anger and mourning on the American right*. The New Press.
- Hohfeld, W. N. (1919). Some fundamental legal conceptions as applied in judicial reasoning. In W. W. Cook (Ed.), *Some fundamental legal conceptions as applied in judicial reasoning and other legal essays, by Wesley Newcomb Hohfeld* (pp. 23–64). Yale University Press.
- Holyoak, K. J., & Cheng, P. W. (1995). Pragmatic reasoning with a point of view. *Thinking & Reasoning, 1*(4), 289–313. <https://doi.org/10.1080/13546789508251504>
- Holyoak, K. J., & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin, 142*(11), 1179–1203. <https://doi.org/10.1037/bul0000075>
- Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General, 128*, 3–31. <https://doi.org/10.1037/0096-3445.128.1.3>
- Hopp, F. R., Fisher, J. T., Cornell, D., Huskey, R., & Weber, R. (2020). The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods, 1*–15. <https://doi.org/10.3758/s13428-020-01433-0>
- Howe, P. D., Mildemberger, M., Marlon, J. R., & Leiserowitz, A. (2015). Geographic variation in opinions on climate change at state and local scales in the USA. *Nature Climate Change, 5*, 596–603. <https://doi.org/10.1038/nclimate2583>
- IPCC. (2007). Summary for Policymakers. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, & H. L. Miller (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment*

- Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
<https://www.ipcc.ch/site/assets/uploads/2018/02/ar4-wg1-spm-1.pdf>.
- Iyer, A., Leach, C. W., & Crosby, F. J. (2003). White Guilt and Racial Compensation: The Benefits and Limits of Self-Focus. *Personality and Social Psychology Bulletin*, 29(1), 117–129. <https://doi.org/10.1177/0146167202238377>
- Jaspers, K. (2009). *The question of German guilt*. Fordham Univ Press.
- Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, 89(6), 899–913.
<https://doi.org/10.1037/0022-3514.89.6.899>
- Kahan, D. M., Landrum, A., Carpenter, K., Helft, L., & Hall Jamieson, K. (2017). Science curiosity and political information processing. *Political Psychology*, 38, 179–199.
<https://doi.org/10.1111/pops.12396>
- Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, 2, 732–735. <https://doi.org/10.1038/nclimate1547>
- Kant, I. (1998). Critique of pure reason (P. Guyer & AW Wood, Trans.). *Cambridge, U. K: Cambridge University Press*.(Original Work Published 1781).
- Kennedy, S. (2020). Few countries on track to meeting climate goals. In *Yale Climate Connections*. <https://yaleclimateconnections.org/2020/01/few-countries-on-track-to-meet-paris-climate-goals/>.
- Koriat, A. (2012). When are two heads better than one and why? *Science*, 336(6079), 360–362.
<https://www.science.org/doi/10.1126/science.1216549>

- Koriat, A. (2018). When reality is out of focus: Can people tell whether their beliefs and judgments are correct or wrong? *Journal of Experimental Psychology: General*, *147*(5), 613. <https://doi.org/10.1037/xge0000397>
- Koriat, A., & Adiv, S. (2016). The self-consistency theory of subjective confidence. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 127–147). Oxford University Press.
- Koriat, A., Adiv-Mashinsky, S., Undorf, M., & Schwarz, N. (2018). The prototypical majority effect under social influence. *Personality and Social Psychology Bulletin*, *44*(5), 670–683. <https://doi.org/10.1177/0146167217744527>
- Kosmitzki, C., & John, O. P. (1993). The implicit use of explicit conceptions of social intelligence. *Personality and Individual Differences*, *15*(1), 11–23. [https://doi.org/10.1016/0191-8869\(93\)90037-4](https://doi.org/10.1016/0191-8869(93)90037-4)
- Kruglanski, A. W. (1990). Lay epistemic theory in social-cognitive psychology. *Psychological Inquiry*, *1*(3), 181–197. https://doi.org/10.1207/s15327965pli0103_1
- Kubin, E., Puryear, C., Schein, C., & Gray, K. (2021). Personal experiences bridge moral and political divides better than facts. *Proceedings of the National Academy of Sciences*, *118*(6). <https://doi.org/10.1073/pnas.2008389118>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*(2), 284–308. <https://doi.org/10.1037/0033-295X.103.2.284>

- Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754–770.
<https://doi.org/10.1016/j.cognition.2008.06.009>
- Lakens, D., & Caldwell, A. R. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*, *4*(1), 251524592095150. <https://doi.org/10.1177/2515245920951503>
- Lee, J., & Holyoak, K. J. (2020). “But he’s my brother”: The impact of family obligation on moral judgments and decisions. *Memory & Cognition*, *48*(1), 158–170.
<https://doi.org/10.1111/j.1559-1816.2009.00517.x>
- Leiserowitz, A., Maibach, E., Rosenthal, S., Kotcher, J., Bergquist, P., Ballew, M., Goldberg, M., Gustafson, A., & Wang, X. (2020). *Climate Change in the American Mind: April 2020*. Yale University and George Mason University.
<https://climatecommunication.yale.edu/publications/climate-change-in-the-american-mind-april-2020/>.
- Leiserowitz, A., Maibach, E., Roser-Renouf, C., Rosenthal, S., Cutler, M., & Kotcher, J. (2018). *Climate change in the American mind: March 2018*. Yale University and George Mason University. <https://climatecommunication.yale.edu/publications/climate-change-american-mind-march-2018/>.
- Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *SAGE Open*, *6*, 1–17.
<https://doi.org/10.1177/2158244016636433>.

- Lewandowsky, S., Cook, J., & Lloyd, E. (2018). The ‘Alice in Wonderland’ mechanics of the rejection of (climate) science: Simulating coherence by conspiracism. *Synthese*, *195*(1), 175–196. <https://doi.org/10.1007/s11229-016-1198-6>
- Lewandowsky, S., Gignac, G. E., & Vaughan, S. (2013). The pivotal role of perceived scientific consensus in acceptance of science. *Nature Climate Change*, *3*, 399–404. <https://doi.org/10.1038/nclimate1720>
- Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Science*, *20*, 748–759. <https://doi.org/10.1016/j.tics.2016.08.001>
- Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A Theory of Blame. *Psychological Inquiry*, *25*(2), 147–186. <https://doi.org/10.1080/1047840X.2014.877340>
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Co.
- Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., & others. (2018). Global warming of 1.5 C. *An IPCC Special Report on the Impacts of Global Warming Of, 1*, 1–9.
- McKibben, B. (2019). *Falter: Has the human game begun to play itself out?* Henry Hold and Co.
- Mitchell, A., Gottfried, J., Kiley, J., & Matsa, K. E. (2014). *Political polarization & media habits*. <https://www.pewresearch.org/journalism/2014/10/21/political-polarization-media-habits/>
- Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods*, *22*(1), 6–27. <https://doi.org/10.1037/met0000086>
- Nagel, T. (1986). *The View From Nowhere*. Oxford University Press.

- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*(2), 291–310.
<https://doi.org/10.1037/0033-295X.108.2.291>
- Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250.
<https://doi.org/10.1037/0022-3514.35.4.250>
- Ochs, E., & Capps, L. (2001). *Living narrative: Creating lives in everyday storytelling*. Harvard University Press.
- Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, *128*(1), 3–72. <https://doi.org/10.1037/0033-2909.128.1.3>
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and persuasion* (pp. 1–24). Springer. https://doi.org/10.1007/978-1-4612-4964-1_1
- Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research*, *41*, 3145–3161.
[https://doi.org/10.1016/S0042-6989\(01\)00173-0](https://doi.org/10.1016/S0042-6989(01)00173-0)
- Powell, D., Merrick, M. A., Lu, H., & Holyoak, K. J. (2016). Causal competition based on generic priors. *Cognitive Psychology*, *86*, 62–86.
<https://doi.org/10.1016/j.cogpsych.2016.02.001>
- Radzik, L. (2001). Collective responsibility and duties to respond. *Social Theory and Practice*, *27*(3), 455–471. <https://doi.org/10.5840/soctheorpract20012735>

- Ranney, M. A., & Clark, D. (2016). Climate change conceptual change: Scientific information can transform attitudes. *Topics in Cognitive Science*, 8, 49–75.
<https://doi.org/10.1111/tops.12187>
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803–814. <https://doi.org/doi:10.1037/0278-7393.21.4.803>.
- Rosentiel, T. (2009). *Partisanship and cable news audiences*. Pew Research Center.
<https://www.pewresearch.org/2009/10/30/partisanship-and-cable-news-audiences/>
- Rossmann, M., & Atkin, M. (2019). *I Work with suicidal farmers. It's becoming too much to bear*. *The New Republic*. <https://newrepublic.com/article/153604/work-suicidal-farmers-its-becoming-much-bear/>.
- Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. (1986). Sequential thought processes in PDP models. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, 2, 3–57.
- Rutchick, A. M., Hamilton, D. L., & Sack, J. D. (2008). Antecedents of entitativity in categorically and dynamically construed groups. *European Journal of Social Psychology*, 38(6), 905–921. <https://doi.org/10.1002/ejsp.555>
- Santos, H. C., Varnum, M. E., & Grossmann, I. (2017). Global increases in individualism. *Psychological Science*, 28(9), 1228–1239. <https://doi.org/10.1177/0956797617700622>
- Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The Church, intensive kinship, and global psychological variation. *Science*, 366(6466), eaau5141.
<https://doi.org/10.1126/science.aau5141>

- Silver, D. (2006). Collective Responsibility, Corporate Responsibility and Moral Taint. *Midwest Studies in Philosophy*, 30(1), 269–278. <https://doi.org/10.1111/j.1475-4975.2006.00139.x>
- Simon, D., Stenstrom, D. M., & Read, S. J. (2015). The coherence effect: Blending cold and hot cognitions. *Journal of Personality and Social Psychology*, 109(3), 369–394. <https://doi.org/10.1037/pspa0000029>
- Smiley, M. (2017). Collective Responsibility. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2017/entries/collective-responsibility/>
- Spellman, B. A., & Holyoak, K. J. (1992). If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology*, 62(6), 913–933. <https://doi.org/10.1037/0022-3514.62.6.913>
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In *Advances in experimental social psychology* (Vol. 21, pp. 261–302). Elsevier.
- Stellar, J. E., & Willer, R. (2018). Unethical and inept? The influence of moral information on perceptions of competence. *Journal of Personality and Social Psychology*, 114(2), 195–210. <https://doi.org/10.1037/pspa0000097>
- Swann, W. B., Jetten, J., Gómez, Á., Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review*, 119(3), 441–456. <https://doi.org/10.1037/a0028589>
- Swim, J. K., & Miller, D. L. (1999). White Guilt: Its Antecedents and Consequences for Attitudes Toward Affirmative Action. *Personality and Social Psychology Bulletin*, 25(4), 500–514. <https://doi.org/10.1177/0146167299025004008>

- Talbert, M. (2019). Moral Responsibility. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University.
<https://plato.stanford.edu/archives/win2019/entries/moral-responsibility/>
- Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*(3), 435–502.
<https://doi.org/10.1017/S0140525X00057046>
- Thagard, P. (2002). *Coherence in thought and action*. MIT press.
- Thagard, P. (2008). *Hot thought: Mechanisms and applications of emotional cognition*. MIT press.
- Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science*, *22*(1), 1–24. [https://doi.org/10.1016/S0364-0213\(99\)80033-0](https://doi.org/10.1016/S0364-0213(99)80033-0)
- Triandis, H. C. (2001). Individualism-Collectivism and Personality. *Journal of Personality*, *69*(6), 907–924. <https://doi.org/10.1111/1467-6494.696169>
- Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology*, *74*(1), 118–128.
- Tropp, L. R., & Wright, S. C. (2001). Ingroup Identification as the Inclusion of Ingroup in the Self. *Personality and Social Psychology Bulletin*, *27*(5), 585–600.
<https://doi.org/10.1177/0146167201275007>
- Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10*(1), 72–81.
<https://doi.org/10.1177/1745691614556679>

- Uhlmann, E. L., Zhu, L. (Lei), Pizarro, D. A., & Bloom, P. (2012). Blood is thicker: Moral spillover effects based on kinship. *Cognition*, *124*(2), 239–243.
<https://doi.org/10.1016/j.cognition.2012.04.010>
- Vignoles, V. L., Owe, E., Becker, M., Smith, P. B., Easterbrook, M. J., Brown, R., González, R., Didier, N., Carrasco, D., Cadena, M. P., Lay, S., Schwartz, S. J., Des Rosiers, S. E., Villamar, J. A., Gavreliuc, A., Zinkeng, M., Kreuzbauer, R., Baguma, P., Martin, M., ... Bond, M. H. (2016). Beyond the ‘east–west’ dichotomy: Global variation in cultural models of selfhood. *Journal of Experimental Psychology: General*, *145*(8), 966–1000.
<https://doi.org/10.1037/xge0000175>
- Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, *18*(3), 247–253. <https://doi.org/10.1111/j.1467-9280.2007.01884.x>
- Wohl, M. J. A., Branscombe, N. R., & Klar, Y. (2006). Collective guilt: Emotional reactions when one’s group has done wrong or been wronged. *European Review of Social Psychology*, *17*(1), 1–37. <https://doi.org/10.1080/10463280600574815>
- Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*(12), 1251–1263.
<https://doi.org/10.1177/01461672982412001>