Impression Formation and Impression Management: Internal and External Determinants

Inauguraldissertation
zur
Erlangung des Doktorgrades
der Humanwissenschaftlichen Fakultät
der Universität zu Köln
nach der Prüfungsordnung vom 10.05.2010
vorgelegt von

Sarah Christina Rom

aus
Köln

Tag der Abgabe: 11.09.2017

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im Dezember 2017 angenommen.

Erstgutachter: Prof. Dr. Christian Unkelbach

Zweitgutachter: Prof. Dr. Wilhelm Hoffmann

Tag der Disputation: 15.12.2017

Erklärung

Kapitel 2 beruht auf folgendem Manuskript:

Critcher, C. R., Dunning, D., & **Rom**, S. C. (2015). Causal trait theories: A new form of person knowledge that explains egocentric pattern projection. *Journal of Personality and Social Psychology*, *108*, 400–416.

Der erste Autor hat die Idee entwickelt, die Datenerhebung überwacht, und die Datenanalyse ausgeführt. Zwei Studien des Manuskripts wurden von mir mitentwickelt und erhoben. Der erste Autor hat das Manuskripts geschrieben, welches durch Beiträge des zweiten Autors und mich profitiert hat.

Kapitel 3 beruht auf folgendem Manuskript:

Rom, S. C., Weiss, A., & Conway, P. (2017). Judging those who judge: Perceivers infer the roles of affect and cognition underpinning others' moral dilemma responses. *Journal of Experimental Social Psychology*, 69, 44-58.

Ich habe die erste Idee entwickelt. Eine Weiterentwicklung der Idee profitierte durch wertvolle Beiträge des zweiten und dritten Autors. Ich habe die Datenerhebung überwacht und die Analyse der Daten ausgeführt. Ich habe das Manuskript geschrieben, welches durch wertvolle Beiträge des zweiten und dritten Autors profitiert hat.

Kapitel 4 beruht auf folgendem Manuskript:

Rom, S. C. & Conway, P. (2018). The strategic moral self: Self-presentation influences moral dilemma responses. *Journal of Experimental Social Psychology*, 74, 24–37.

Ich habe die erste Idee entwickelt. Eine Weiterentwicklung der Idee profitierte durch wertvolle Beiträge des zweiten Autors. Ich habe die Datenerhebung überwacht und die Analyse der Daten ausgeführt. Ich habe das Manuskript geschrieben welches durch wertvolle Beiträge des zweiten Autors profitiert hat.

Sarah Rom, Köln, September 2017

Summary

When forming impressions of other people or when managing our own impressions, people are influenced by internal and external information. After a literature overview in Chapter 1, Chapter 2 investigates how people form impressions based on internal influences (e.g., self). Results showed that as people try to understand themselves, they develop *causal trait theories*, theories that explain how their standing on one trait is caused by their standing on another trait. Six studies showed that people's theories about the self guided their understanding about how traits relate in people in general. Chapter 3 investigates how people form impressions based on external factors (e.g., a person's overt judgment). Across six studies I found that participant's inferred traits from a target's moral dilemma judgment. A target who rejected harm was perceived as relatively warm and moral, but rather incompetent, whereas a target that accepted harm to maximize outcomes was perceived as relatively competent, but rather cold and immoral. Chapter 4 examined how both, internal (e.g., meta-perceptions) and external factors (e.g., context/situation) influence how people manage their impressions. Results demonstrated that people have accurate meta-perceptions, that is, they knew how they were perceived by others following a moral dilemma judgment. This meta-insight influenced in turn how people managed their impressions. Depending on which judgment was favored in a given situation people publically shifted their moral dilemma responses, although privately their judgments remained unaffected. These results show, that people have access to higher order processing that causally contributes to moral decision-making. Chapter 5 discusses limitations, open questions, and implications.

Keywords: impression formation, impression management, moral dilemmas, social perception, meta-perceptions

Zusammenfassung

Wenn wir einen Eindruck über andere Menschen bilden oder wenn wir unseren Eindruck auf andere steuern, werden wir von internalen und externalen Informationen beeinflusst. Nach einem literarischen Überblick in Kapitel 1, zeigt Kapitel 2 empirisch, wie die Eindrucksbildung von internalen Einflussfaktoren beeinflusst werden kann am Beispiel von Selbstwahrnehmung. Ergebnisse von sechs Studien zeigten, dass Theorien, die Probanden darüber hatten, wie Persönlichkeitseigenschaften bei ihnen selbst kausal zusammenhängen, beeinflussten, wie sie dachten, dass Eigenschaften bei anderen Personen zusammenhängen. Kapitel 3 untersucht, wie die Eindrucksbildung von externalen Faktoren beeinflusst werden kann am Beispiel von moralischen Entscheidungen einer anderen Person. Wie Ergebnisse von sechs Studien zeigen, inferierten Probanden Eigenschaften basierend auf den moralischen Dilemma-Entscheidungen einer Zielperson. Eine Person, die die charakteristisch deontologische („Das Töten von Leben ist unter allen Umständen falsch“) Entscheidung traf, wurde als relativ warm und inkompetent wahrgenommen, wohingegen eine Person, die die charakteristisch utilitaristische („Das Töten von Leben ist gerechtfertigt, wenn das Allgemeinwohl dadurch gesteigert wird“) Entscheidung traf, als relativ kompetent und kalt wahrgenommen wurde. Kapitel 4 untersuchte, wie sowohl internale (*hier:* Meta-Wahrnehmung) als auch externale (*hier:* Kontext/Situation) Informationen beeinflussen, wie andere Leute die Eindrücke, die sie auf andere machen, steuern. Ergebnisse von sieben Studien zeigten, dass Personen sich bewusst waren, wie sie von anderen Personen aufgrund ihrer moralischen Entscheidungen wahrgenommen wurden. Diese Meta-Einsicht beeinflusste wiederrum, wie Menschen die Eindrücke, die sie auf andere machen, steuern. In der Öffentlichkeit gaben Personen das moralische Urteil, welches im jeweiligen Kontext den besseren Eindruck machte, im Privaten blieb ihr Urteil unbeeinflusst. Kapitel 5 diskutiert Einschränkungen, offene Fragen, und Implikationen der empirischen Befunde.

Schlagwörter: Eindrucksbildung, Eindruckssteuerung, Selbstwahrnehmung, Meta-Wahrnehmung, Moralische Dilemma

Acknowledgements

This dissertation could not have been completed without great support over the years. I wish to thank the following people:

First and most of all, thank you, Christian. I still remember the first time we met, I was thinking of applying to Cologne and you told me about the position. This was way before I met any of my amazing colleagues, way before SoCCCo even exhisted, way before I knew what my dissertation topic would be. I was immediately sold, maybe due to your rare combination of being extremely warm and extremely competent, which proved to be true in every meeting we had these past four years. I thank fate that you were looking for a PhD student right when I was looking for a position. Thank you for accepting me as a grad student, for inspiring me, giving me endless ideas on what to study, and thank you for caring so deeply about me.

Wilhelm, thank you for your mentorship and friendship. I am very grateful we met over melted chocolate pears and started to collaborate not much later. Thank you for sitting down with me to write (be it a paper or rebuttal), for helping me practice my talks, for advising me when I needed advice, for encouraging me when I felt discouraged, and also, for inviting me to come to New York. This was one of the very best conference and symposium experiences I had and I will cherish it forever. Last but not least, thank you for offering me some of my most intellectually exciting interactions over the past four years, not only during labor but also during leisure time.

My unofficial third advisor and friend, Paul, thank you for all the meetings and discussions, for your willingness to teach me, and for having hard conversations with me. Thank you for caring so deeply about my research and career. It has been an epic, windy, rocky road, and this dissertation would not have been possible without you. As we are now divided in time and space, I know our friendship will continue.

Alex, thank you for always being there, but especially for being there right in the beginning when I needed it the most. Thank you for teaching me how to code (you remember those nights in the office) and for always believing in me when I did not believe in me. Being able to share research in everyday life and also being able to share hotel rooms at conferences with you have always been my highlights.

Hans, you made me feel welcome to the Unkelbach family right from day one: From eating at the Mensa, to spending New Years, to going on holidays, to partying at weddings or conferences (and every year you manage to be the first in our flat when we celebrate Carneval). I cannot be anything else but impressed by your intelligence and brilliance, which has probably helped and inspired me more than you even know. I was so lucky to call you and Alex my colleagues.

My office mate Fabia, after 3 years sharing an office I can say I have probably spent the most time over these past four years with you. Thank you for your kindness, for being pleasant and so considerate. It is much nicer to come to work when you look forward to a nice office and person to come to every day.

Marita, thank you for giving me the warmest welcome I have ever been given anywhere in my life. I will never forget the first e-mail I received from you. You left to soon but I will always remember you, our conversations, and your good nature.

Laura, our messenger conversations could fill several books, I am happy we went through the Ph.D together, and that our friendship has nothing but grown closer over these past four years. I am beyond excited to see what lies ahead of us as we finish our Ph.ds and both start a new chapter of our lives.

Thanks to the rest of the SoCCCo family for making conferences, pre-conference holidays, Mensa gatherings, carnival festivities, SoCCCo retreats the high points of my Ph.D.

Table of contents

Chapter 1 – Introduction

Forming impressions of others and managing our own impression are pervasive tasks in everyday life. Whenever we encounter someone for the first time we instantly infer personality traits based on limited information we get about a person (Ambady, Rosenthal, 1993; Uleman, Sarbay, & Gonzales, 2008). For example, on a first date we infer how likable they are, when interviewing someone for a job we infer how competent they are, and when listening to someone's talk we infer how wordy they are. Naturally, in all these instances, we are keen on leaving a favorable impression as well (Reis & Gruzen, 1976; von Baeyer, Sherk, & Zanna, 1981; Leary & Kowalski, 1990; Leary, 1995). No matter if we are the ones to judge or to be judged, perceptions and behavior can be influenced by internal (e.g., self-, and meta-perceptions) or external (e.g., a person's behavior, a situation) factors. For example, when going on a date, we may infer how much we like someone depending on internal information we have about ourselves or external information we observe from someone else's behavior (e.g., Atkinson & Feather, 1966). That is, we may think: "He made a huge effort to plan the perfect date, I only do so if I like someone very much, therefore he must like me", or "He made a huge effort to plan the perfect date, therefore he must like me". In the first instance, internal information we have about ourselves influences our perception of them. In the second instance, external information we get from another person's observable behavior influences our social perception.

In the Chapters that follow I focus on impression formation and impression management from these two perspectives. To do so, in Chapter 2 I investigate how internal information (e.g., self-perception) influence impression formation. I hereby investigate how *causal trait theories*, theories generated to explain the relationship between two traits in the self are used to explain people in general. In Chapter 3 I investigate how external information (e.g., a person's judgment) influence impression formation. I hereby investigate how people perceive others on the dimensions' warmth, competence and morality based on their overt moral dilemma judgment. Needless to say, there exist a lot of other external determinants, and a person's moral dilemma judgments is just one example. In Chapter 4 I investigate how internal (e.g. meta-perception) and external (e.g., situation) information influence how people manage their impressions following a moral dilemma judgment. Before I show the empirical evidence I will first discuss the relevant literature.

1.1 Internal Influences of Person Perception

A swath of evidence supports the notion that how we see other people is intertwined with how we see ourselves (Sherif & Hovland, 1961; Alicke, Dunning, & Krueger, 2005; Dunning, 2002). Whenever we are trying to predict a person's behavior, judge their performance or evaluate their personality the self plays a fundamental role in the way we perceive and understand other people (Anderson & Ross, 1984; Pronin, Kruger, Savitsky, & Ross; 2008). For example, people who spent much time in volunteering rate a target's morality lower than people who hardly engage in volunteering (Dunning & Hayes, 1996). Further, people make more extreme and confident judgments of others on traits they see as self-descriptive (Fong & Markus, 1982; Lambert & Wedell, 1991). By defining social traits in self-serving ways, people are able to place themselves in the most favorable light.

People also assume that traits they possess will be more common in other people as well, a phenomenon called *attributive projection* (Goldings, 1954; Katz & Allport, 1931; Kruger & Stanke, 2001). However, recent research has found that the influence of the self on social judgment does not stop there. People do not only project traits but also the way how two traits are related, a phenomenon called *egocentric pattern projection* (Critcher & Dunning, 2009). In egocentric pattern projection the correlation of two traits observed in the self is consistent with how two traits are correlated in other people. If Jane, a creative extrovert learns that Joe is also an extrovert she will assume that Joe must also be creative. Thus, Jane expects that the patterning of traits in others will recapitulate the patterning of traits in the self. Speaking in analysis of variance terms, attributive projection predicts two main effects, whereas pattern projection predicts and interaction. Pattern projection was even replicated when using a fictitious personality questionnaire: after learning that they were front-brained and V-dominant, as opposed to back-brained and Z-dominant, participants assumed that two traits were patterned in other people in the same way as observed in the self. These and a plethora of other results demonstrate that the self plays a special role in the way we perceive other people, and that our understanding of others is colored by our own self-understanding. Instead of looking across exemplars in a category and trying to develop a theory of why certain characteristics would co-occur with others, people can simply look at the self and develop a theory about why certain traits co-occur. Although I am unaware of past research that has documented that process, there is precedent for certain steps in this proposed chain. First, people are quick at developing causal theories on the spot based on limited information, even when it would make more sense to wait for more information before drawing conclusions (Chater & Oaksford, 2005; Keil, 2006). Second, there is evidence that people will generalize learning about a single exemplar to other exemplars of the category (Risen, Gilovich, & Dunning, 2007; see also Lopez, Gelman, Guthiel, & Smith, 1992). Although these examples do not describe the precise internal process proposed here, they do hint that people are quick to draw explanatory conclusions, even with little supporting evidence, which can then influence the way people think about others. Building on these and related findings, the present dissertation explains the self's special role for the emergence of egocentric pattern projection, arguing that egocentric pattern projection emerges as a byproduct of trying to understand why traits relate as they do in the self. After people notice a correlation between two traits in themselves they form a *causal trait theory* to explain their patterning, and this theory forms the foundation for the expectation in how these two traits relate in others. Causal trait theories allow people to go beyond single trait information they know about someone. For example, Jane, the creative extrovert, may explain how spending time with others helps her fuel her creativity and therefore expects the traits in others to be patterned the same way they are patterned in the self. Pattern projection emerges when people start applying self-specific theories to understand people in general. To sum up, Chapter 2 of the present dissertation demonstrates how causal trait theories people form to understand the self are projected in other people.

1.2 External Influences of Person Perception

People not only form impressions of others' traits (Carlston & Skowronski, 1994) based on internal information, but also external information such as behavior, actions, or overt judgments—impressions which guide subsequent information processing, decision-making (Ambady & Rosenthal, 1992) and evaluative valence (Schneid, Carlston, & Skowronski, 2015). Such inferences go beyond surface-level features: they pertain to the personality behind overt judgments and behavior.

One such instance in which this may prove useful is the case of a person's moral dilemma judgment. In the hijacked airplane dilemma, for example, a passenger jet has been hijacked by terrorists, and is now heading towards a densely populated urban center. Is it acceptable to shoot this plane down—including the innocent civilians on board—in order to prevent it from wreaking widespread carnage? Imagine a student named Brad agrees and says harm is acceptable. What impression would his decision leave upon you? How warm, competent, or moral would you judge Brad to be?

People typically pay particular attention to other people's moral decisions (DeScioli & Kurzban, 2008), as morality is perceived as the core of the self (Strohminger & Nichols, 2014) and thus especially diagnostic of personality (Goodwin, Piazza, & Rozin, 2014). Moreover, perceptions of others' morally-relevant judgments and behavior often provoke powerful responses ranging from elevation (Schnall, Roper, & Fessler, 2010) and gratitude (McCullough, Kilpatrick, Emmons, & Larson, 2001), to gossip (Feinberg, Willer, Stellar, & Keltner), social distancing (Skitka, Bauman, & Sargis, 2005), punishment (Fehr, Fischbacher, & Gächter, 2002) and ostracism (Feinberg, Willer & Schulz, 2014). For example, people consider intentions when assigning blame, independent of outcomes (Cushman, 2008), people make character inferences from other's morally relevant decision making (Critcher, Inbar, & Pizarro, 2013), and they form negative impressions of others who technically did nothing wrong if their actions suggest bad moral character (Inbar, Pizarro, & Cushman, 2012; Pizarro & Tannenbaum, 2011). Haidt (2001) even argued that moral judgments are social in nature: speakers try to articulate their moral standpoint in order to reach consensus with other people. Therefore, moral judgments communicate important information about the speaker.

Given the social importance of morality and people's tendency to gain insight into the psychology behind others' moral judgments and behavior, it seems plausible that people pay attention to others' moral dilemma judgments, draw inferences regarding the processes that led to such judgments, and use these inferences to inform perceptions of others' personality.

I propose that people infer how affect and cognition underpin others' moral dilemma judgments, and use this information to draw inferences about others' personality and social decisions. Accordingly, targets who make characteristically deontological judgments (e.g. *causing harm is inappropriate regardless of outcomes*) should be perceived as relatively warm, because perceivers infer that such judgments are driven by affective reactions to harm. Conversely, targets who make characteristically utilitarian judgments (e.g. *causing harm is appropriate when it maximizes overall outcomes*) should be perceived relatively more competent, because perceivers infer that such judgments are driven by cognitive deliberation.

If, as I predict people are attuned to the social impact of dilemma judgments—this raises the question of whether people have meta-insight into how their dilemma judgments make them appear in the eyes of others (meta-perception). If so, this gives plausibility to the possibility that people shift dilemma responses depending on which trait is valued in a situation (Chapter 4).

1.3 Internal and External Influences of Impression Management

Chapter 4 focuses on internal (e.g., meta-perceptions) and external (e.g., context) influences on impression management. I argue that just as self-perception influences the way we see other people (i.e., impression formation), meta-perception—estimates of how others might view us—may influence the way how we want to be seen by others (i.e., impression management).

Despite the fact that real-world judgments often take place in the presence of, or are communicated to, other people (Haidt, 2001; Hofmann, Wisneski, Brandt, & Skitka, 2014) moral dilemma research has mostly ignored the social context of dilemma judgments. Indeed, some work suggests a role for motivated processing in moral judgments (Uhlmann, Pizarro, Tannenbaum, & Ditto, 2009; Liu & Ditto, 2014), and that dilemma judgments are sensitive to social influence (Kundu & Cummins, 2012). People are motivated to *appear* moral (Batson & Thompson, 2001; Batson, 2008) and to make and share moral judgments of other people (Feinberg, Willer, Stellar, & Keltner, 2012). Besides, a swath of evidence on impression management has shown that people tailor their public images in various domains to the perceived values and preferences of significant others (Reis & Gruzen, 1976; von Baeyer, Sherk, & Zanna, 1981; Leary & Kowalski, 1990; Leary, 1995). I propose that if people are aware of third party judgments, people might modify their moral dilemma decision depending on whether warmth or competence is valued more in a given situation. In contexts where affect is valued, people may be more likely to reject causing harm, whereas in contexts where cognition is valued, people may be more likely to accept causing harm to maximize outcomes. If people are strategic about managing their impressions when forming dilemma judgments, participants should publicly choose the decision in a way that lets them appear most favorably in a given situation—private judgments, however, should remain unaffected. These results would be the first to suggest that moral dilemma judgments arise out of more than just intrapsychic cognitive and affective processes; complex social considerations causally shape moral dilemma decision-making.

1.4 The current research

The present dissertation looks at impression formation and impression management from two perspectives. In Chapter 2 I examine how internal determinants in the form of causal trait theories influence the way how we see other people. More specifically, I argue that causal trait theories represent a case in which people, as trying to make sense of themselves, end up projecting their own self-perception in others.

In Chapter 3 I investigate the influence of external determinants (e.g., a person's moral dilemma judgments) on person perception. That is, I propose that how we see other people can be influenced by another person's observable action (e.g., judgment). More specifically, I argue that people make trait inferences based on a target's moral dilemma decision. A target who rejects causing harm will be perceived as relatively warm and moral, but rather incompetent, a target who accepts causing harm to maximize outcomes will be perceived as relatively competent, but rather cold and immoral.

If people have meta-insight into how their dilemma judgments make them appear in the eyes of others (meta-perception), people may be able to shift dilemma responses depending on which trait is valued in a situation. Thus, in Chapter 4 I investigate internal (e.g., meta-perceptions) and external (e.g., situation) influences on impression management. I propose that people have accurate meta-perceptions following a moral dilemma judgment, which in turn influence how people manage their impressions. Therefore, I predict that people strategically modify their dilemma judgments to leave a favorable impression, either by selecting an answer that is favored in a given situation, or by framing a judgment in a self-serving way.

Chapter 2 – Causal Trait Theories: A New Form of Person Knowledge That Explains Egocentric Pattern Projection

Abstract

We argue that person representations include not merely piecemeal lists of traits, but also *causal trait theories*—explanations for why a person's standing on one trait causes or is caused by the person's standing on other traits. We find people hold such theories about people, but especially about the self (Studies 1a-1c). Further, these causal theories resolve the puzzle of *egocentric pattern projection*—the tendency for people to assume that traits correlate in the population in the same way they align in the self. Causal trait theories—created to explain trait co-occurrence in a single person—are exported to guide one's implicit personality theories about people in general. Supporting this analysis, when people have a causal trait theory to explain why two traits relate in a person—either the self or someone else—they assume more of a correlation between those two traits in the broader social world (Study 2). Pattern projection was found to be egocentric for two reasons. First, causal trait theories are more numerous for the self; when people were prompted to generate causal trait theories about someone else, they pattern projected more from that person (Study 3). Second, causal trait theories about the self are mostly created to explain why two abstract traits coexist in a single person, whereas causal trait theories about others are also generated in response to observing a person display two traits in a given context. A yoked design showed that causal trait theories produce pattern projection only when they drawn on behavioral information across multiple contexts (and thus are a theory about personality) instead of when they draw on behavior from a single context (Study 4).

KEYWORDS: pattern projection, self, causal thinking, implicit personality theories, narratives, person perception, egocentrism

Causal Trait Theories:

A New Form of Person Knowledge That Explains Egocentric Pattern Projection

A person's perspective on his or her social world is typically framed by the self. Whether in taking another's perspective, predicting the opinions of others, or evaluating people one encounters, the self's own perspective (Epley, Keysar, Van Boven, & Gilovich, 2004; Epley, Morewedge, & Keysar, 2004), characteristics (Dunning, Meyerowitz, & Holzberg, 1989), and standing (Dunning & Cohen, 1992; Dunning & Hayes, 1996) influence such judgments. This egocentrism exists, in part, to maintain a person's sense of self-worth (e.g., Beauregard & Dunning, 1998), but it also permeates social views for other reasons. The self's own egocentric perspective is effortlessly brought to mind, and adjusting away from it is effortful (Epley et al., 2004). Furthermore, in a social environment that is sparse on information, relying on self-knowledge may be a reasonable heuristic for understanding others (Dawes, 1989).

In this paper, we seek to explain a recently-documented means by which self-perception colors social perception, *egocentric pattern projection* (Critcher & Dunning, 2009). To do so, it will be necessary to offer and test a new account of why our understanding of others is contaminated by our understanding of ourselves. We posit a new type of person knowledge—one that goes beyond mere facts about a person (e.g., "I am health-conscious and protective") to incorporate theories of why one aspect gives rise to or causes another (e.g., "My being protective leads me to be health-conscious because…"). Ultimately, we will argue that as a byproduct of trying to make sense of themselves, people end up coloring their impressions of others.

2.1 Egocentric Pattern Projection

In displaying *egocentric pattern projection*, people seem to use how two traits relate in the self to infer whether the two traits are positively or negatively correlated in other people. For example, if Jens sees himself as egalitarian and emotional, Jens will expect egalitarian people to be emotional and nonegalitarians to be less so. If Jens, instead, sees himself as egalitarian and not very emotional, he will assume that egalitarians will be unemotional, but that nonegalitarians will be more emotive.

Across five studies, Critcher and Dunning (2009) provided consistent support for this pattern of aligning traits in others like one does in the self, showing that people's implicit personality theories (IPTs)—beliefs about how personality traits tend to be configured in people in general—tended to recapitulate the way traits were patterned in the self. Critcher and Dunning also distinguished this type of projection from its simpler cousin, *attributive projection*, in which people merely assume that individual traits they possess are more common in other people (Goldings, 1954; Holmes, 1981; Judd, Kenny, & Krosnick, 1983; Katz & Allport, 1931; Krueger & Stanke, 2001; Ross, Greene, & House, 1977). For example, with attributive projection, an egalitarian and emotional Jens would presume that other people are commonly emotional and egalitarian than he would if he did not possess those traits, but would not draw inferences about the relationship between the two traits—that is, whether they wax and wane in others in tandem.

Critcher and Dunning (2009) additionally showed that the self played a causal role in pattern projection. When a (fictitious) personality inventory informed participants they were front-brained and V-dominant, as opposed to back-brained and Z-dominant, participants assumed the two traits were correlated in a way consistent with the patterns observed in the self. Participants concluded that other people tended to be either front-brained and V-dominant or back-brained and Z-dominant. However, when given someone else's personality feedback, there was no similar jump to assume a correlation between brain-type and variation of dominance.

2.2 Causal Trait Theories as the Missing Link

Critcher and Dunning (2009) documented pattern projection as a novel phenomenon, but offered no empirical data as to why it arose or why it was egocentric. The present paper seeks to fill this void by focusing on a new type of person knowledge, *causal trait theor*ies.

Psychologists have long appreciated that when we understand a person, we know more than a mere list of trait descriptors (McAdams, 1985, 2001). As we come to know someone better, we progress from simple trait ascriptions to a better understanding of their personal strivings and motivations, to ultimately developing a coherent narrative that achieves coherence, meaning, and purpose by weaving together events in the person's past, present, and anticipated future (Adler & McAdams, 2007; McAdams, 1995; Pals, 2006).

We agree that person representations are richer than mere traits, but we emphasize that even trait knowledge can take a more sophisticated form than a simple listing about a person's standing on various characteristics. Consider how a research participant studied by Park (1986) described another: "She is wealthy and egotistical, which makes for great fashion sense and good looks." This statement not only describes four piecemeal features of the social target, but offers a causal theory about how they relate: [(wealthy + egotistical) → (fashionable + attractive)]. This suggests impressions not only contain listings of traits, but also theories about how such attributes are causally related (Murphy & Medin, 1985).

The idea of causal trait theories has origins in Asch (1946), who noted that impressions of others—based on a list of traits—are different from the mere "sum" of those traits. In this sense, he recognized that traits in others are not interpreted in isolation, but have implications for how other traits in that person should be interpreted or inferred. Thus, the calmness displayed by a warm person is qualitatively distinct from the calmness displayed by a cold person. Asch also focused on the importance of the order in which another's traits are learned, finding that earlier-learned information constrains the way later-learned information is understood.

Despite our different empirical focus, our interest in causal trait theories is foreshadowed in Asch's (1946) theoretical approach, in which he argued that people "try to get at the root of personality", that this means "the traits are perceived in relation to each other," and that such impressions comment on “processes between the traits each of which has a cognitive content” (p. 259). Whereas Asch used these ideas to justify why trait-based understandings cannot be studied in isolation, we draw on these ideas as pointing toward an important type of perception in its own right. That is, causal trait theories reflect explanations of how traits are influencing one another—the “processes between the traits.” Although previous research has not set out to document causal trait theories directly, previous work has found that people have little difficulty generating *ad hoc* theories on demand (McNorgan, Kotack, Meegan, & McRae, 2007), even about seemingly contradictory evidence (Asch & Zukier, 1984). This gave us confidence that *causal trait theories* may be a pervasive but overlooked aspect of person representations.

How might causal trait theories help to explain pattern projection? Historically, there has been debate about whether implicit personality theories (IPTs) are represented as mere correlations or associations among traits, as a multi-dimensional factor space onto which trait relationships can be mapped, or as “person types” (Kim & Rosenberg, 1980; Rosenberg, 1976; Anderson & Sedikides, 1991). We instead propose that implicit personality theories are, in fact, theories—i.e., rich explanations that go beyond mere correlation coefficients, factor loadings, or trait clusters (see also Sedikides & Anderson, 1994). As such, they contain a rich representation of how traits are causally related to each other. This assumption is consistent with recent empirically-supported theorizing that people learn diagnostic relationships between features (i.e., whether the presence of X signals the presence of Y) by determining whether the evidence is consistent with a causal connection between the two (Meder, Mayrhofer, & Waldmann, 2014).

Research in cognitive psychology has uncovered the important role of explanatory or causal theories in perceived correlations (Chapman & Chapman, 1967; Kunda, Miller, & Claire, 1990; McNorgan et al., 2007; Murphy & Wisniewsky, 1989). Ahn, Marsh, Luhman, and Lee (2002) illustrated this principle by showing that people are often unaware of actual, observable correlations when they are difficult to explain. For example, most people are aware of the correlation between how close to water a bird lives and the probability that fish is part of a bird's diet (Ahn et al., 2002). In contrast, far fewer people realize that among shirts, there is a correlation between the presence of buttons and the length of the sleeves. The former correlation lends itself to a simple causal narrative (e.g., "If a bird wants to eat fish, it behooves it to live near the ocean"), whereas there is no obvious causal theory to explain the latter (positive) correlation.

Ahn et al. (2002) noted that their research left open the question of whether "people explicitly notice correlations because they can explain them, or [whether] people impose explanations after they explicitly notice correlations" (p. 115). Our account of pattern projection proposes a mix of these two ideas. We suggest that instead of looking across exemplars in a category and trying to develop a theory of why certain characteristics would co-occur with others, people will also look to a single exemplar—usually, but not always the self—and develop a theory about why certain characteristics (e.g., traits) co-occur in that sample of one. Supportive of this idea, people seem quite comfortable and ready to develop causal theories on the basis of one-shot learning, even when it would seem much more reasonable to remain agnostic until observing a broader array of data (Chater & Oaksford, 2005; Keil, 2006). In addition, people generalize conclusions they draw about a single exemplar (e.g., a minority group member) to other members of the category (Risen, Gilovich, & Dunning, 2007; see also Lopez, Gelman, Guthiel, & Smith, 1992).

Of course, people can have causal trait theories about any person—either themselves or someone else. But Critcher and Dunning (2009) found that pattern projection is egocentric—that is, stronger for patterns of traits in the self than for patterns of traits in well-known others. There are two independent ways that our account could explain such egocentrism. We test both possibilities.

The quantity hypothesis. A first possibility, the *quantity hypothesis*, is that people are more likely to generate causal trait theories to understand the self as opposed to someone else. If pattern projection emerges as a byproduct of generating causal trait theories to explain a single person, then people should pattern project more from targets about which they have generated more theories. This account does not see causal trait theories about the self as *special,* just more numerous. Thus, it predicts that people should pattern project from others as well when they have generated causal trait theories to explain them.

Of course, it would be naïve to predict that people never engage in similar theorizing about others, but we contend that such theories may be narrower in scope and more simplistic in structure. A number of previous findings support this possibility. Although people compose "person models" to explain others, these models tend to be structured around a central trait, with other information linked to this core concept (Park, DeKray, & Kraus, 1994). This leaves room only for causal theories that include the core concept. And even when representations of others include many traits, factor analyses indicate that representations of others are organized in a more simplistic and redundant manner than are understandings of the self (Beer & Watson, 2008; Borkenau & Liebler, 1994). The structure of other representations is more likely to follow a simple "evaluative narrative" (e.g., "She's a jerk") that does not necessitate a rich causal structure (Hampson, 1998). In total, self-knowledge is more nuanced, comprehensive, and complex. Causal trait theories may provide the glue to unify this disparate self-knowledge.

The breadth hypothesis. *A* second possibility, *the breadth hypothesis,* is that causal trait theories generated to explain the self are different in the breadth of their origin than theories about others. Causal trait theories can originate from people seeking to explain why two abstract qualities or traits exist within the same person by drawing upon information from *multiple contexts* (e.g., "Does the fact that I am so conscientious, like when I'm at work, explain why I am often so quiet, like when I'm at home?"). Also, causal trait theories can originate from people seeking to explain why a person behaves as he or she does in a *single context* (e.g., “Was Mary so attentive at the party because she was feeling not very confident about her cooking?”) We suggest that when causal trait theories take the former form—i.e., they draw on information from multiple contexts—they are more likely to be pattern projected. After all, these are theories about personality—unifying explanations of why one person may display different behaviors in different contexts. In contrast, theories of the latter variety are explanations of behavior in a context, meaning they should not be as easily exported to become general theories of human personality.

If causal trait theories that describe the self are more likely to be *multiple context* theories than are causal trait theories about others, then this could be a second source of egocentrism in pattern projection. Although ultimately this is an empirical question (which we tackle), there are a few reasons to think this would be true. The self is, tautologically, with itself in more contexts than it is with others. As such, the self has more cross-context information for it to drawn upon as it reflects on itself. Furthermore, the self has direct access to its own intentions, but not those of others. This means that in any single context, there will be more of a demand to make sense of someone else’s co-occurring behaviors instead of one’s own. We test these assumptions and whether they account for the egocentric nature of pattern projection.

2.3 Overview of the Studies

In sum, we propose that causal trait theories are an overlooked aspect of person knowledge and a key construct that will help to resolve the lingering mystery of why egocentric pattern projection emerges. Studies 1a to 1c introduced three distinct methods to test for the prevalence of causal trait theories, ultimately assessing whether such theories are more numerous and accessible about the self than about others. Study 2 tested whether causal trait theories explain egocentric pattern projection. Studies 3 and 4 provided experimental tests of the quantity and breadth hypotheses by testing whether people begin to pattern project from others once prompted to think about others in the style, and with the type of information, that characterizes the way people tend to think about the self. Study 3, in a test of the quantity hypothesis, tested whether prompting people to generate causal trait theories (vs. memorize trait information) about others encourages pattern projection from them (as the quantity hypothesis would predict). Study 4, in a test of the breadth hypothesis, tested whether participants who received behavioral information about yoked participants that spanned multiple contexts (thereby matching the informational origin of causal trait theories for the self), versus that came from a single context, generated causal trait theories that encouraged relatively more pattern projection from the yoked others.

2.4 Study 1a

Study 1a was designed to examine the prediction that people held a greater number of causal trait theories for the self than for others. Participants were asked to create a trait theory map, either of themselves or their freshman year roommate. We chose roommates as the comparison other for two reasons: 1) roommates have been used as a "familiar other" in prior research (Prentice, 1990), and 2) Critcher and Dunning (2009) repeatedly established that college students pattern project more from themselves than from their freshman-year roommate.

2.4.1 Method

Participants and Design. Two hundred eight undergraduates at Cornell University participated in exchange for $5 or extra course credit. Participants were randomly assigned to draw a causal trait theory map to describe themselves (*self* condition) or their freshman year roommate (*other* condition).

Procedure. All participants began by rating themselves or their freshman year roommates[1] on sixteen personality traits: *bashful*, *considerate*, *cunning*, *dependent*,

[1] If participants had more than one freshman-year roommate, they were asked to choose the roommate whose bed was closest to the participant's own. If participants did not have a roommate, they were asked to consider the person to whom they lived closest.

extravagant, *generous*, *happy-go-lucky*, *idealistic*, *opportunistic*, *persistent*, *prideful*, *prudent*, *reserved*, *resigned*, *skeptical, wordy*. We included this step because we did not want differences between causal trait theory maps to emerge only because trait knowledge about the self was more accessible.

Participants were then given 16 index cards, each representing one of the sixteen traits. All were then told that people sometimes construct theories to explain people, explanations that link together different aspects of one's personality in a causal story. To facilitate thinking about causal trait theories, participants were first asked to look through the cards and form clusters of traits for which a theory could be offered to explain why all those traits co-existed within the person. The instructions explained that each cluster had to have at least two traits in it, and that participants need not use all 16 cards. Because each trait appeared on exactly one card, the same trait could not appear in multiple clusters. Although there are several interesting, measurable features of these clusters (e.g., how many traits are part of the clusters created, how many "theory clusters" did participants create altogether), this step was largely a prelude to the next stage, in which we had people draw out more complete causal theory maps.

In the next task, participants were told that they would draw a more complete causal trait map, indicating the ways in which specific personality traits influenced other traits, or how two traits were influenced by some third-variable aspect of personality. Two examples were offered to illustrate the difference between these types of theories. One was a direct causal link with, "In me, I am creative *because* I am not very extroverted." Participants represented such direct causal link by drawing a directional arrow from one trait to another.   The other was a third-variable causal link, "My desire to grow up to be a successful artist leads me to further develop my creative abilities and to spend a lot of time on solitary activities that are not very extroverted." Participants represented a third-variable link by connecting two traits with a line, and then drawing an arrow that pointed at the line (see Figure 1). Our primary motivation in assessing links of both types (direct or third-variable) was to understand whether one type was obviously more prevalent than the other, and thereby guide our focus in future studies.

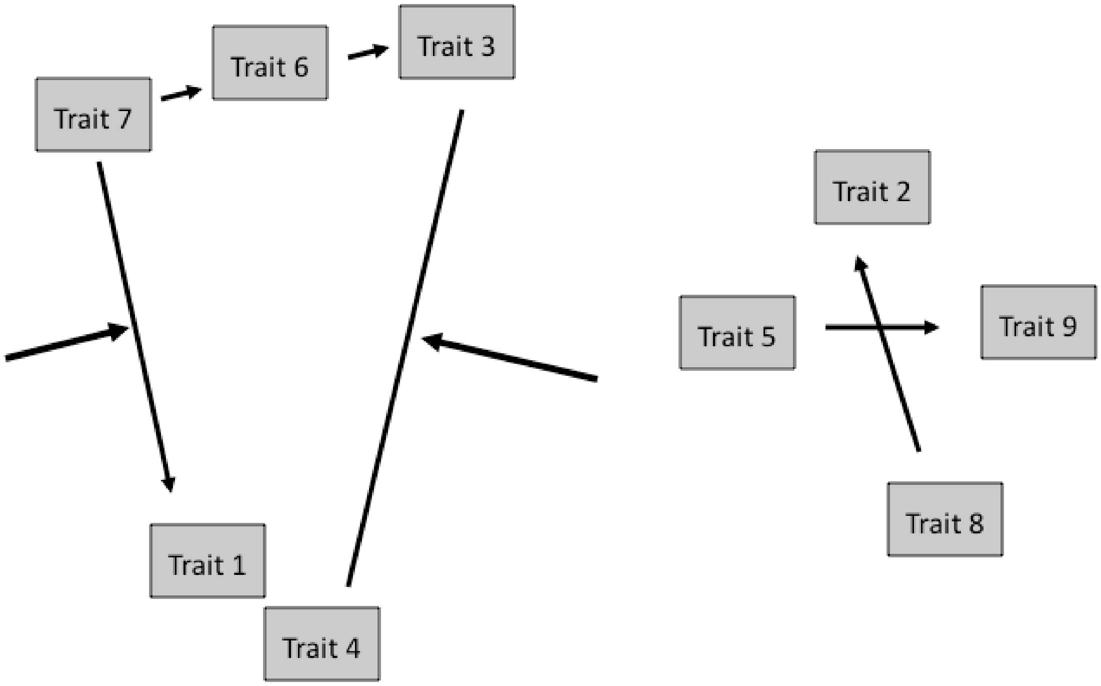


Figure 1. Example causal trait theory map from Study 1a. This maps depicts 5 directional theories and 2 third-variable theories. Note that two traits can be connected by theories of both types (e.g., Traits 1 and 7).

2.4.2 Results and Discussion

By every metric, the trait theory maps of the self were more comprehensive and contained more causal connection than those of roommates. As seen in Table 1, when describing the self, participants created a larger number of clusters than when describing an other, $t(202) = 2.22$, $p = .03$, $d = .31$. Furthermore, they included more of the 16 traits in their own clusters than in those describing someone else, t(182.26)[2] = 2.81, p = .01, d = .40. In addition, participants saw more direct causal relationships in the self than they did in someone else, t(206) = 2.17, p = .03, d = .30. Although third-variable theories were relatively rare in characterizing either target, such theories were also more numerous in maps of the self than those of the roommate t(195.17) = 2.38, p = .02, d = .33.

It is notable the extent to which directional theories were much more numerous than third-variable theories in the theories participants had about the self, paired t(105) = 20.39, p < .001, d = 1.98, and about the other, paired t(101) = 22.32, p < .001, d = 2.21. As such, we only measure directional theories in future studies, given that the rarity of third-variable theories make them poor candidates for explaining pattern projection. Furthermore, we see Study 1a as an especially conservative test of our hypotheses. That is, even if participants did not already have well-formed causal trait theories about the other, they may have tried to create them in the moment. Study 1b explores this possibility further.

Attribute	Self	Other
Clusters	4.65 (0.99)	4.31 (1.21)
Traits used	13.40 (1.89)	12.52 (2.55)
Directional Causal Theories	10.89 (5.03)	9.53 (3.88)
Third-Variable Causal Theories	1.27 (1.58)	0.81 (1.19)

Table 1. Features of Causal Trait Theory Maps Describing the Self or an Other (Study 1a).

[2] When independent-sample t-tests include a non-integer degree of freedom, this reflects a correction due to a homoscedasticity violation. The degrees of freedom in the multi-level models were calculated using the Satterthwaite approximation.

Each mean is followed by the corresponding standard deviation.

2.5 Study 1b

Did participants' causal trait theory maps reflect pre-existing representations, or did they merely reflect people's constructions once they were prompted to describe them? Study 1b addressed this question by assessing whether causal trait theories about the self are not only more numerous but more accessible in memory. That is, if such theories already exist, they should be more rapidly reportable (Park, 1986; Prentice, 1990). If instead participants in Study 1a had more theories about the self because they took more time trying to construct them in the moment, then it would take longer for participants to report this information about the self than about someone else.

Participants in Study 1b were asked fifty-five yes/no questions about whether they had a causal trait theory to explain why two traits were related in the self. They also answered the same questions concerning their causal trait narratives of their freshman year roommate. We had two central predictions. First, we expected that even with this modified measurement technique, people would again report more causal trait theories for the self than for an other. Second, we expected that these causal trait theories for the self would be more accessible than the theories for the roommate. That is, people should be faster to indicate whether they have a theory to explain the self than a theory to explain the roommate.

2.5.1 Method

Participants. Participants were 41 undergraduates from Cornell University. In exchange for their participation, participants received $5 or course credit.

Procedure. As in Study 1a, participants began by rating themselves and their freshman year roommates, with order counterbalanced across participants, on eleven personality traits: *bashful, considerate, dependable, happy-go-lucky, idealistic, persistent, prideful, reserved,* *resigned, skeptical, wordy*. Next, participants answered 55 questions about their causal trait theories for the self as well as 55 about the causal trait theories for their freshman year roommate. Each question took the form, “Does how SKEPTICAL you are [your roommate is] cause how PRIDEFUL you are [your roommate is]?” For each trait pair, which trait was the possible antecedent vs. consequent trait (in this case, skeptical vs. prideful, respectively) was held constant. Responses were coded dichotomously (yes = 1; no = 0).[3] The time participants took to respond—from the moment the question appeared on screen until the point that the participant depressed one of the two response keys—was recorded, in milliseconds. The order of responding about the self versus the roommate was counterbalanced across participants[4].

2.5.2 Results and Discussion

We first tested whether participants reported more causal trait theories for the self than for their roommates. We submitted the total number of “yes” responses to a mixed-model ANOVA, with the counterbalancing order variable as a between-subjects variable and target (self or other) as a within-subjects variable. As predicted, participants’ self-narratives were more numerous (M = 28.3, SE = 1.42) than their causal narratives about their roommates (M = 24.5, SE = 1.44), $F(1, 39) = 13.93$, $p < .001$, $\eta_p^2 = .26$. Also, participants more quickly answered causal questions about themselves (M = 4.02s, SE = .24s) than about their freshman year roommates, (M = 4.36s, SE = .29s), $F(1, 39) = 6.43$, $p = .02$, $\eta_p^2 = .14$. This accessibility difference was

[3] We used a smaller sample size in Studies 1b and 1c (compared to Study 1a) because each participant reported causal trait theories for both the self and the roommate (instead of one or the other). This increased our power because it allowed us to control for individual differences in the tendency to report having causal trait theories. Highlighting the gains that the fully within-subjects design offered, we observed a strong correlation between how many causal trait theories participants reported having about the self and their roommate: $r(39) = .75$, $p < .001$ (Study 1b) and $r(71) = .76$, $p < .001$ (Study 1c).

[4] There was no evidence of significant skew in the latencies to the self-theory (z = 1.61, p = .11) or the roommate-theory questions (z = 1.45, p = .15), so all analyses on the reaction times are performed on the untransformed means.

equally strong regardless of whether participants were indicating that a particular trait pair was or was not in their causal narrative for their roommate or for themselves, $F < 1$.

These accessibility findings are particularly helpful in that they help to speak against an artifactual account of Study 1a that participants were merely willing to spend more time, in the moment, trying to generate or "fish for" causal trait theories about the self than someone else. If so, participants would have been slower, not faster, to report on causal trait theories about the self.

2.6 Study 1c

Study 1c introduced a third, more conservative method to measure the presence of causal trait theories. Participants were given a trait, and then asked whether they could generate another trait in themselves (or their freshman year roommate) that explained their (or their roommate's) standing on the first trait. If they indicated they could, they had to list what that causal trait was. In this way, participants were more accountable when they indicated they had a causal trait theory: Research participants tend to report less information in their self-representations when they have to generate the content themselves as opposed to merely indicate whether certain knowledge is in these representations (Dunkel & Anthis, 2001). Thus, the present method would give us more confidence that self-other differences reflected differences in person representations as opposed to differences in a willingness to endorse that items are part of one's self-representation.

As a secondary goal, we tested whether participants reported having causal trait theories in a circumstance that should, logically, predict their presence: whether the person (i.e., the self or the other) was seen as highly consistent (as opposed to variable) on the trait across situations. People should be more likely to explain a consistent trait by appealing to something about the person (i.e., by forming a causal trait theory); in response to cross-context variability, situational explanations may become more likely. If causal trait theories were more numerous for consistent traits, we would have further confidence that the causal trait theory measure was valid, and did not merely reflect more self-theories because of a bias toward indicating that one's self-knowledge is more thorough than it actually is.

2.6.1 Method

Participants. Participants were 73 undergraduates from the University of California, Berkeley. In exchange for their participation, participants received course credit.

Procedure. Participants saw the 11 personality traits used in Study 1b. As before, participants rated themselves and their roommates on the traits. Next, they completed two measures in a counterbalanced order:

Causal trait theories. Participants were told that they would be asked to indicate their theories of why they (and their freshman year roommates) had certain traits. We explained that they may have a theory that "how much you have or display…a given trait is caused or influenced by some other trait you possess." As an example, we described Mary, a woman whose *unkind* nature might be explained by her being *ambitious*: "Perhaps Mary (rightly or wrongly) believes she does not need other people to get ahead, so her ambitiousness leads her to be ruthlessly unkind to others." For each trait, participants were to indicate whether another trait in themselves (or their roommate) "influences or causes how much you have (or your freshman-year roommate has) the trait." We then explained that, "If you cannot think of such a trait, type 'none' in the blank." The traits appeared in a random order.

Cross-situational consistency. We explained to participants the difference between showing consistency or variability in how much one shows a trait. An example described that a person might be moderately jittery from situation to situation (*consistent*), or may be quite jittery in some situations but not at all jittery in others (*variable*). For each trait, participants indicated (dichotomously) whether it existed fairly consistently or with variability in the self (or in the roommate).

2.6.2 Results and Discussion

Given our interest in multiple levels of analysis we used multi-level modeling. We tested whether participants were more likely to indicate a causal trait when considering why the self possesses traits than when considering why their freshman year roommate does. We began with one Level-1 variable, *target* (self = +1, roommate = -1), nested within participant. We also defined *order*, a Level-2 variable meant to differentiate participants who indicated their causal trait theories about the self or the roommate first. The target X order interaction term accounted for variance that was merely attributable to the order in which the target measures were completed. Finally, we included a random effect of *trait*, because some traits prompted more theories than did others. This analysis revealed that participants were more likely to identify causal traits for the self (73.74%) than for their freshman-year roommate (67.02%), $t(71) = 3.19$, $p = .002$, semi-partial $R^2 = .13$. Thus, three studies using different measures with different strengths converge on the conclusion that people have more causal trait theories to explain themselves than to explain others.

Supporting our secondary goal, people were more likely to have a causal trait theory to explain consistent traits (74.44%) than variable traits (64.28%), $t(71.43) = 4.34$, $p < .001$, semi-partial $R^2 = .21$. This relationship was true for theories about the self and one's roommate alike, $t < 1$[5]. Thus, two aspects of this study—the fact that participants had to identify the causally antecedent trait, as well as the presence of the systematic negative relationship between the

[5] Note that because, if anything, trait consistency is seen to be lower in the self (38.27%) than in the roommate (46.21%), $t(71) = 3.25$, $p = .002$, semi-partial $R^2 = .13$, trait consistency likely suppresses, but certainly does not explain, the self's advantage in number of causal trait theories versus the freshman-year roommate (Monson et al., 1980).

presence of a causal trait theory and the reported stability of the trait in the target—lends support to the validity of the causal trait theory measure.

2.7 Study 2

Having provided convergent evidence that people hold causal trait theories, especially for the self, we more directly examined their role in explaining egocentric pattern projection. Participants provided trait ratings about their own and their freshman year roommate's personalities, and provided judgments about how pairs of personality traits are correlated in people in general (i.e., implicit personality theories). We expected to replicate Critcher and Dunning (2009) by uncovering evidence of egocentric pattern projection. That is, we expected that implicit personality theories would relate to how traits were patterned in the self (pattern projection), but less so how traits were patterned in the roommate (egocentric).

Participants also indicated whether they had a causal trait theory to explain why each of the 55 trait pairs was related in the self or in the roommate, using a measure similar to that used in Study 1b. If causal trait theories underlie pattern projection, then people should show stronger pattern projection for those trait pairs whose co-occurrence is explained with a causal trait theory. If the greater number of causal trait theories to explain the self versus the roommate accounts, at least in part, for pattern projection's egocentrism (the quantity hypothesis), then we should expect causal trait theories to be more numerous in the self (as in Studies 1a-1c), but also for the theory's presence to moderate the degree of pattern projection from both the self as well as the roommate. If it is not merely the number, but the nature of causal trait theories for the self (vs. someone else) that explains the egocentric nature of pattern projection (consistent with, but not necessarily supportive of the breadth hypothesis), then the presence of a causal trait theory for the self should predict more pattern projection than the presence of a causal trait theory for an other.

2.7.1 Method

Participants and design. Participants were 213 undergraduates at the University of California, Berkeley, who participated in exchange for course credit or $15.

Procedure. All participants provided trait judgments of themselves and their freshman-year roommates, indicated whether they had causal trait theories to explain trait co-occurrences in the self and in the roommate, and made judgments from which their implicit personality theories could be induced. Participants completed their trait judgments and implicit personality theories in a counterbalanced order. Either 30 minutes before or 30 minutes after completing these, participants indicated whether they had a causal trait theory to explain how each of 55 trait pairs co-occurred in the self and in the roommate. Self and roommate judgments were also made in a counterbalanced order:

Causal trait theories. Participants answered a total of 110 dichotomous questions: 55 about the self, and 55 about their freshman-year roommate. Each question was of the same form: "Does how RESIGNED you are [your freshman-year roommate is] cause how CONSIDERATE you are [your roommate is]?" Participants responded by indicating yes ('Y') or no ('N').

Trait judgments. Participants indicated their own standing, and their freshman-year roommate's standing, in a counterbalanced order, on each of the 11 traits. The 11-point scale was anchored at 1 (*not at all*) and 11 (*extremely*).

Implicit personality theories. Participants answered one question for each of the 55 trait pairs: "If all you knew about a person was that he or she was more _______ than average, it is what percent likely that s/he would also be more __________ than average?" To make sure people understood the logic of the scale, we noted that al responses should be between 0% and 100%. To establish 50% as a neutral midpoint, the experimenter noted that "If knowing someone is more [the first trait] than average gives you no information about whether the person is more [the second trait] than average, you would indicate 50%." The order in which the

IPT was measured always matched the order in which the causal trait theory was measured. That is, the example causal trait theory measure provided above would be paired with an IPT measure asking how likely a person who is more resigned than average would also be more considerate than average.

2.7.2 Results

First, we attempted to replicate our earlier findings that people have more causal trait theories in understanding the self than someone else. Second, we attempted to replicate Critcher and Dunning's (2009) finding that people pattern project more from the self than from someone else. Third, we tested whether the greater number of theories people have about the self explains why people pattern project more from the self (the quantity hypothesis). Fourth, we tested whether theories people have to explain the self are more likely to prompt pattern projection than theories people have to explain someone else (consistent with the breadth hypothesis).

Do people have more causal trait theories about the self? We submitted participants' responses to the causal trait theory measure to a 2 (target: self or roommate) X 55 (trait pair) mixed-model ANOVA. Conceptually replicating the earlier results, participants had more causal trait theories to explain themselves (M = 26.7, SE = 0.7) than to explain their freshman-year roommate (M = 25.0, SE = 0.7), F(1, 211) = 14.37, p < .001, η_p^2 = .06.

Do people pattern project more from the self than from their roommate? Next, we attempted to replicate Critcher and Dunning's (2009) findings that people's implicit personality theories recapitulate patterns observed in the self more than someone else. First, we defined two Level-1 variables that were centered before being entered into all analyses: *self-difference* and *roommate-difference*. For any given pair of traits i and j, the variables reflected the absolute value of the difference between the trait judgments for the self or the roommate on those two traits, respectively. Pattern projection is observed when the degree to which two traits co-occur similarly [dissimilarly] in a target predicts beliefs that the two traits correlate positively [negatively] in the general population.[6]

We constructed a random-slope, random-intercept model predicting participants' implicit personality theories, in which self-difference and roommate-difference were nested within each trait pair. In this way, we could explain whether individual differences in implicit personality theories for a specific trait pair could be traced to differences in people's perceptions of their own (and their freshman-year roommate's) personality. The random intercept essentially controls for differences between trait pairs in how much they are perceived as correlated, but we also included *participant* as a random effect to control for individual variability in seeing traits, in general, as more positively or negatively correlated.

There was evidence of pattern projection both from the self and from the roommate. That is, the greater the difference in any two traits in the self, B = 1.12, *SE* = 0.10, or in the freshman-year roommate, B = 0.69, *SE* = 0.10, the more people held the implicit personality theory that they two traits were negatively correlated in people in general, *t*s = 11.02 and 6.62, *p*s < .001, respectively. To test whether pattern projection was egocentric, we ran an additional model that compared the relative influence of the two predictors in predicting implicit personality theories—i.e., whether the two betas just reported were significantly different. They were, *t*(19,816.84) = 2.75, *p* = .01, semi-partial R^2 = .0004. In short, people pattern project more from themselves than they do from another sample of one—i.e., their roommate.

Do causal trait theories explain egocentric pattern projection? We extended our last model by first introducing two more Level-1 variables: *self-theory* and *roommate-theory*. Each variable was coded +1 if, for that particular trait pair for that particular participant, the participant indicated having a causal trait theory to explain the self (self-theory) or their

[6] Pattern projection is reflected by negative betas, but for ease of interpretation, all such betas, in this and all studies, have been reversed so that positive values reflect pattern projection.

freshman-year roommate (roommate-theory). The same variables were coded -1 if participants reported not having such a theory. We tested whether causal trait theories encourage pattern projection from both the self and the roommate (the quantity hypothesis). We then tested whether causal trait theories were more likely to encourage pattern projection from the self than from the roommate (consistent with the breadth hypothesis). Note these hypotheses are not mutually exclusive.

Do people pattern project trait relationships for which they have causal trait theories? Supporting the quantity hypothesis that causal trait theories give rise to pattern projection, both self-difference X self-theory and roommate-difference X roommate theory interaction terms were significant (see Figure 2). More specifically, people pattern projected from the self more for trait pairs for which they had causal trait theories to explain the trait co-occurrence in the self, B = 0.38, *SE* = 0.08, *t*(7,209.12) = 4.54, *p* < .001, semi-partial R^2 = .0028. Turning to simple effects, when participants had a causal trait theory to explain why two traits co-occurred as they did in the self, they pattern projected strongly, B = 1.43, *SE* = 0.16, *t*(10,999.16) = 8.85, *p* < .001, semi-partial R^2 = .0071. But when participants failed to have a causal trait theory to explain the co-occurrence in the self, pattern projection was significantly weaker, B = 0.67, *SE* = 0.16, *t*(22,912.47) = 4.23, *p* < .001, semi-partial R^2 = .0008.[7]

[7] In another study not reported here, we found that whether people pattern projected from a given trait pair depended on whether they have a directional causal theory (e.g., “Does how CONSIDERATE you are cause how WORDY you are?”) to explain why the traits were related, *t*(5,509.81) = 3.29, *p* = .001, semi-partial R^2 = .0020, but not on whether they had a third-variable theory (e.g., “Does another aspect of your personality [for example, a goal you have or a trait you possess] help explain both how CONSIDERATE you are and how WORDY you are?”) that explained the traits, *t*(7,382.46) = -0.09, *ns*. Thus, not only are third-variable theories not particularly numerous in person representations (Study 1a), they do not appear to play an important role in understanding pattern projection. We speculate that this comes from a difference in “exportability” of the two theories. A theory that explains why X causes or is caused by Y (a direct causal theory) can be applied relatively unconditionally. In contrast, a theory that both X and Y are influenced by Z (third variable theory) is more straightforward to apply when one knows another’s standing on Z. Regardless, this suggests that pattern projection emerges not merely from two traits being connected as part of a broader narrative, but requires that traits be directly linked.

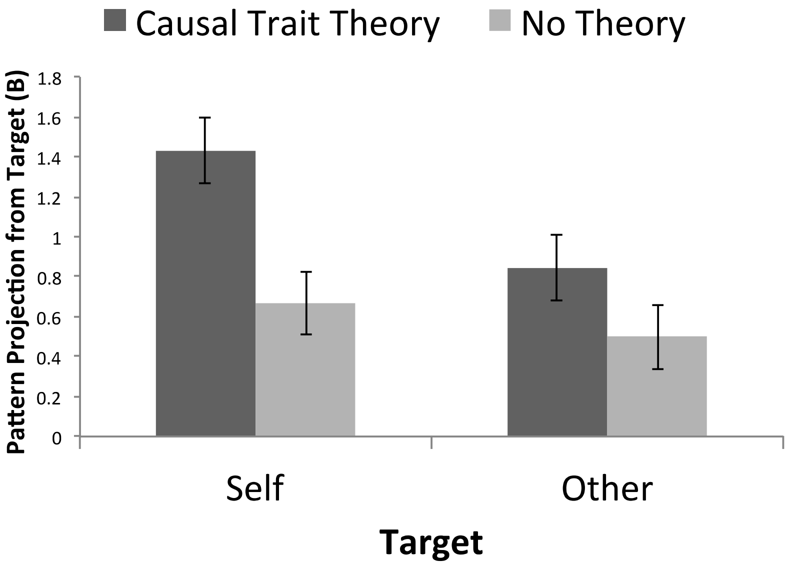


Figure 2. Pattern projection from a target (self or other) for trait pairs for which participants did or did not report having a causal trait theory about that target (Study 2). The significant difference in bars within each cluster reflects the Self-difference X Self-theory and Roommate-difference X Roommate-theory interaction terms—i.e., support for the quantity hypothesis). The significant difference between the two darker (causal trait theory) bars is consistent with the breadth hypothesis. The error bars reflect + 1 SE of the estimate of the beta that corresponds to pattern projection. Because implicit personality theories were measured somewhat differently in Study 2 as opposed to Studies 3-4, it is not meaningful to compare betas across those studies.

Participants also pattern projected from their freshman-year roommate more for trait pairs for which they had causal trait theories to explain trait co-occurrence in the roommate, B = 0.17, *SE* = 0.08, *t*(6,582.27) = 2.16, *p* = .03, semi-partial R^2 = .0007. That is, when participants had a causal trait theory to explain trait co-occurrence in the roommate, they pattern projected from this other, B = 0.85, *SE* = 0.17, *t*(24,637.40) = 5.05, *p* < .001, semi-partial R^2 = .0010. But when participants did not hold such a causal trait theory, pattern projection from the roommate was weaker, B = 0.50, *SE* = 0.16, *t*(35,895.44) = 3.14, *p* = .002, semi-partial R^2 = .0003, but still significant. The fact that pattern projection—even in the absence of causal trait theories—was still significant, both from the self and from the roommate, suggests that multiple mechanisms may give rise to pattern projection.

Do causal trait theories about the self produce more pattern projection than do causal trait theories about another? We next tested whether there is something special—as the breadth hypothesis would predict—about the causal trait theories about the self that predict more pattern projection. We first tested whether the Self-theory X Self-difference interaction term was stronger than the Roommate-theory X Roommate-difference interaction term. The difference was marginally significant, $t(6{,}190.65) = 1.77$, $p = .08$, semi-partial $R^2 = .0005$.

Stronger support for our account was found once we moved on to the planned comparisons. In particular, when participants had a causal trait theory to explain both the self and a roommate, they pattern projected more from the self, $t(13{,}354.13) = 2.56$, $p = .01$, semi-partial $R^2 = .0005$. But in the absence of any causal trait theories, there was no greater pattern projection from the self than from the roommate, $t < 1$, semi-partial $R^2 < .0001$. This latter finding shows that causal trait theories fully explain the *egocentric* nature of pattern projection.

2.7.3 Discussion

Study 2 supported our contention that causal trait theories play a crucial role in producing pattern projection and explaining its egocentric nature. First, people were more likely to pattern project from the self or from someone else when they had a causal trait theory to explain a given trait patterning in that person (consistent with the quantity hypothesis). But second, and consistent with the breadth hypothesis, when people had a causal trait theory to explain the self, that translated into significantly stronger pattern projection than did a causal trait theory to explain the roommate. In combination, this suggests that part, but not all, of the egocentric nature of pattern projection is explained by the greater number of causal trait theories people hold about the self versus someone else. However, an additional part must be explained by some feature of causal trait theories about the self, which are more likely to generalize to influence implicit personality theories compared to similar theories about the roommate.

In Studies 1a-1c, we were sensitive to the possibility that participants may not have had preexisting causal trait theories, but may have been constructing them only once suggested by the measures themselves. Although we provided evidence that spoke against that possibility, we took advantage of our counterbalancing in Study 2 to address this question in an additional way. Recall that some participants indicated whether or not they had causal trait theories about the self and the roommate *before* stating their implicit personality theories, whereas for other participants the order of these measures was reversed. If asking people to report on their causal trait theories caused people to create theories they did not already have, then we should see a stronger link between the presence of theories and pattern projection when the causal trait theory measure preceded the implicit personality theories compared to when causal trait theories were measured later. Contradicting this possibility, the tendency for pattern projection to be stronger from trait pairs about which participants had a causal trait theory was not stronger still when causal trait theories were measured before implicit personality theories. That is, there was no further moderation by the order manipulation in explaining when pattern projection emerged from the self, B = 0.02, *SE* = 0.08, $t < 1$, or from the roommate, B = 0.08, *SE* = 0.08, $t(10{,}365.02) = 1.03$, $p > .30$, semi-partial $R^2 = .0001$. If the measures themselves were prompting the creation of causal trait theories, then implicit personality theories should have been less tethered to causal trait theories when causal trait theories were measured at the end (when any newly-created theories could no longer influence the implicit personality theories).

One limitation of Study 2 is that the results are correlational. Is it possible that people observe trait patternings in the world, explain them, and then have causal trait theories to make sense of those same patternings in themselves and their roommate? This reverse-causality argument is made unlikely by Critcher and Dunning's (2009) experimental studies that showed that self-perceptions lead to implicit personality theories. Furthermore, this alternative explanation cannot easily account for why causal trait theories more strongly predict pattern projection for the self than for the roommate.

A final concern is whether there may be certain trait pairs—perhaps because they share semantic overlap (e.g., bashful and reserved)—that are more likely to co-occur in the self, more likely to be explained by causal trait theories, and also more likely to be perceived as correlated. This is essentially a third-variable concern. Two features of our design and analyses helped rule out this alternative. First, because all analyses look at the influence of the self while controlling for the influence of the roommate, one would have to explain why a feature of a trait pair like semantic overlap would appear in one's self-ratings and self-theories but not in one's roommate-ratings and roommate-theories. Second, because all multi-level models were nested within trait pair (and thus explain variation between participants while holding the trait pair constant), this essentially prevents general differences among trait pairs from driving effects.

2.8 Study 3

Study 3 was designed to provide *causal* support for the first proposed mechanism underlying egocentric pattern projection—that having a causal trait theory to explain why two traits relate in a person leads people to export that theory to explain people in general (the quantity hypothesis). In Study 3, we presented participants with trait information about three novel targets. Some participants constructed causal trait theories to explain why traits related as they were said to in those targets. Other participants processed the information about these targets in a more piecemeal fashion that did not involve theorizing. After processing the target information in one of the two ways, participants stated their implicit personality theories.

We expected that participants who generated causal trait theories to explain a specific other would begin to pattern project from that person. Of course, our account does not predict that pattern projection emerges from the mere *attempt* to create a causal trait theory to explain someone else, but instead from the successful creation of the theory. This predicts a more nuanced hypothesis that *only* to the extent that a participant reports success in generating a causal trait theory about the target should that attempt elevate pattern projection from that target.

2.8.1 Method

Participants and design. Participants were 405 undergraduates at Cornell University who completed the experiment in exchange for extra course credit. Participants were randomly assigned to one of four conditions in a 2 (processing task: causal theory or control) X 2 (target version) full-factorial design.

Materials and procedure. Participants were first informed that they would receive information about three different targets. This information would comprise four sentences, each conveying trait-relevant information. As seen in Table 2, each sentence was of the form, “Person X is [very; very much NOT] trait Y.” After each sentence appeared on the screen, 45 seconds elapsed before the next sentence would appear. Even after the next sentence would appear, the prior sentences were still visible. During the 45-second period, participants were to engage in one of two processing tasks, depending on their condition.

Target 1 (Version A; Version B)

Person 1 is (not at all; very) generous.

Person 1 is not at all cunning.

Person 1 is very resigned.

Person 1 is very dependent.

Target 2 (Version A; Version B)

Person 2 is (not at all; very) happy-go-lucky.

Person 2 is very bashful.

Person 2 is not at all prideful.

Person 2 is (not at all; very) idealistic.

Target 3 (Version A; Version B)

Person 3 is (very; not at all) skeptical.

Person 3 is (not at all; very) prudent.

Person 3 is very opportunistic.

Person 3 is very wordy.

Table 2. Both Versions of the Three Social Targets Presented in Study 3. Trait information is listed in the order in which it was presented.

Targets. We chose twelve traits that we had used in prior studies. All participants had indicated their own standing on each of the traits (all listed in Table 2) on a web-based pretest completed at least 24 hours before coming to the lab. These ratings were made on eleven-point scales anchored at 1(*not at all me*) and 11 (*completely me*). These traits were randomly grouped into three groups of four traits. The four traits in each group would form the basis for a description of a novel target. We then constructed two versions of each person by randomly determining, for each target, whether the target was described as “very much” having the trait, or “not at all (being the opposite of)” the trait.[8] To minimize the likelihood that two contrasting traits would be nonsensically paired within the same person (e.g., very bashful but not at all reserved), we added the constraint that the four traits used to describe a person had to be fairly uncorrelated. (We used trait ratings from past studies to confirm that the absolute value of each correlation was less than .20.)

The two versions of each of the three targets are described in Table 2. What is less important than the level of each trait in each target is how each pair of traits relates in each target. For example, even though Target 3’s skepticism and prudence differ by version, the two traits are “negatively correlated” in each version and thus do not constitute a trait pair of interest. Across the three targets, 11 of the 18 observed trait relationships differed between the two targets. If participants are pattern projecting from a target, then they should infer more of a positive correlation between two traits when the traits relate similarly in the target (“very much” – “very much” or “not at all” – “not at all”). They should infer more of a negative correlation when the traits exist dissimilarly in the target (“very much” – “not at all” or “not at all” – “very much”).

Causal trait theory condition. The instructions in the causal theory condition prompted participants to generate a causal trait theory explaining how the traits all influenced each other to give rise to a single, coherent individual:

[8] We followed the precedent of Critcher and Dunning (2009) in saying that a person was “not at all” a trait instead of trying to find a word to characterize the opposite of a trait. This afforded three advantages. First, this facilitated our measurement of implicit personality theories, for we could use a single trait label to refer to each trait dimension. Second, this permitted a more efficient presentation of materials, for we did not have to teach participants which traits they should assume to be the exact opposites of which traits. Third, we were not limited by having to lean only on traits that had clear opposites.

> "Your task will be to incorporate each new piece of information you learn into a coherent picture of the person. You want to try to link together individual traits to understand how they influence or affect each other, why they fit together as they do in the same person."

We then provided an example causal trait theory that could explain why a person was both very extroverted and very creative. It was emphasized to participants that they should try to analyze and type for the full 45 sec. Note that we gave plenty of time to participants to create these theories because such theories are more content-rich than the simple relationship "trait A causes trait B." Instead, theories involve a fuller explanation about *why* such a relationship emerges.

Control condition. For the control task, it was important that participants still focus on information about the target, but not on how or why the traits co-occurred in the target. Accordingly, control participants were asked to elaborate on what it meant for the target to possess each of his or her traits. Thus, when each new sentence appeared, instead of spending 45 seconds trying to generate theories to connect the newly presented trait to the other traits, the participant spent 45 seconds elaborating on what the trait meant. "For example, if you learned a person was 'very much extroverted,' you might type that the person is 'a sociable, affable kind of person, interested in socializing, not at all aloof or shy, warm, gregarious…'" It was emphasized that participants were to generate these descriptions only about the most recently presented piece of information. This was stressed so participants would not think their task was to synthesize across the traits and describe what the person as a whole was like.

After participants completed the full 3-min processing task for each target, participants were asked how difficult it was to successfully complete the processing task for that target by pressing 1 (not at all difficult), 2 (a little difficult), 3 (somewhat difficult), or 4 (very difficult). After seeing all three targets, participants stated their implicit personality theories for all eighteen possible trait pairs, even though only eleven of these trait pairs would allow us to assess whether participants were pattern projecting from the targets. We measured implicit personality theories using Critcher and Dunning's (2009) three-judgment method—assessing *p*(trait 1), *p*(trait 2), and *p*(trait 1 | trait 2). The IPT was derived using the following linear expression: *p*(trait 2) * [*p*(trait 1 | trait 2) – *p*(trait 1)]. Higher numbers reflected a greater perceived correlation.

At the end of the experiment, participants encountered a surprise recognition task. Participants were presented with the twelve traits that had been associated with the three targets. They had to indicate whether the target in question "very much" or "very much did NOT" have the trait. In this way, we could assess whether any tendency to pattern project differently by condition could actually be attributable to a superior explicit memory for the information about the target instead of the act of theorizing about the target.

2.8.2 Results and Discussion

We tested whether those assigned to generate a causal trait theory of someone else would then pattern project more from that person. First, we created a variable called *patterning*. This variable differentiated whether a specific pair of traits, as seen by a specific participant, was patterned in the target in a way that implied a positive correlation (+1: very X"—"very Y" or "not at all X"—"not at all Y") or a negative correlation (-1: "very X"—"not at all Y" or "not at all X"—"very Y"). Thus, a positive effect of patterning on implicit personality theories would reflect pattern projection. Second, we define the variable *processing task*, which differentiated participants who were prompted to generate causal trait theories (+1) versus process the traits in a piecemeal fashion (-1). Third, given previous research indicates that people tend to pattern project from the self, we used participants' pretest ratings of themselves to create absolute value difference scores for all relevant traits pairs (i.e., | Self rating on trait i – Self rating on trait j|).

We constructed a multi-level model to assess our main hypotheses. Patterning, processing task, (self-reported) difficulty (of the processing task), and the self-difference score were nested within trait pair in a random-slope, random-intercept model. This permitted the effects of the predictors to vary by trait pair (random-slope), but also allowed the general implicit personality theory for each trait pair to vary (random-intercept). In addition to the higher-order interaction terms, we included the categorical variable *participant*, which corrected for differences between participants in the extent to which they tended to see trait pairs as more or less correlated.

Overall, participants pattern projected from the targets they learned about, B = 70.49, *SE* = 25.87, *t*(9,895.30) = 2.73, *p* = .01, semi-partial R^2 = .0011. But also, the degree of pattern projection depended on the processing task condition to which they had been assigned, B = 36.24, *SE* = 18.79, *t*(3,503.21) = 1.93, *p* = .05, semi-partial R^2 = .0011. Participants prompted to generate causal trait theories to explain a specific target began to pattern project from that target, B = 106.83, *SE* = 31.33, *t* = 3.41, *p* = .002. Participants in the control condition who were prompted to analyze the trait-based information in a piecemeal fashion did not pattern project from the target, B = 34.19, *SE* = 32.74, *t* = 1.04, *p* > .29.

But note our central hypothesis is more nuanced. That is, we do not predict that people will pattern project from someone else merely because they have *attempted* to generate a causal trait theory about that person. Instead, people should be especially likely to pattern project from the target when they find they are able to generate such a causal trait theory. A Patterning X Processing Task X Difficulty[9] interaction revealed that those who were more successful in generating causal narratives about a target showed greatest evidence of pattern projection from

[9] People found it marginally more difficult to complete the control task (*M* = 2.70, *SE* = .037) than the causal narrative task (*M* = 2.60, *SE* = .036), *F*(1, 399) = 3.23, *p* = .07, η_p^2 = .01. As such, we standardized the difficulty ratings separately by processing task condition before running the model.

the targets, B = -62.87, *SE* = 19.17, *t*(3,553.04) = 3.28, *p* = .001, semi-partial R^2 = .0030. For those who generated causal trait theories about the target, they pattern projected from that target when they found it relatively easy (-1 SD) to generate this narrative, *t* = 4.95 *p* < .001, but not when they found it difficult (+1 SD) to do so, *t* < 1. Participants in the control condition did not pattern project from the target, regardless of whether the found it easy or difficult, *t*s < 1.13, *p*s > .26, to describe the traits the target possessed (see Figure 3).

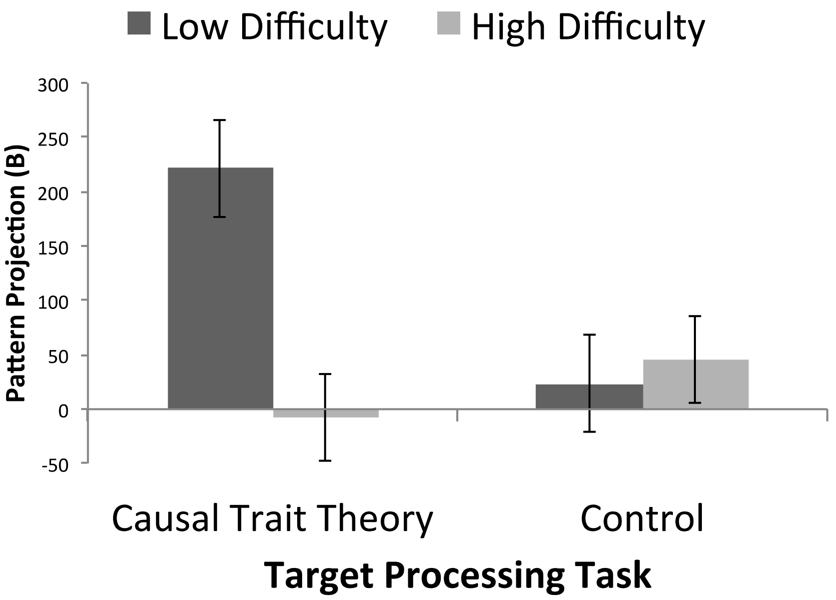


Figure 3. Pattern projection from the novel social targets as a function of processing task condition and the difficulty participants had with the processing task (Study 3). High and Low difficulty is predicted at + 1 standard deviation from the mean level of self-reported difficulty in the respective processing task condition. The error bars reflect + 1 SE of the estimate of the beta that corresponds to pattern projection. Because implicit personality theories were measured somewhat differently in Study 2 as opposed to Studies 3-4, it is not meaningful to compare betas across those studies.

These findings still leave open the question of whether writing a causal trait theory about a person influenced the way people thought about people in general (i.e., their implicit personality theories) for a different reason—enhanced memory for details of the target (Hamilton, Katz, & Leirer, 1980). People better remember two contiguous stimuli when they are brought into the same perceptual unit, such as when they are seen as cause and effect (Asch, 1946). And as these previously-documented findings foreshadow, participants in the causal trait theory condition did indeed have a better memory for the targets' standing along the traits (M = 10.19, SD = 2.06) than did those in the control condition (M = 9.20, SD = 2.40), t(386.53) = 4.46, p < .001, d = .44. There was no evidence, however, that superior memory for the trait information was the mediator responsible for the impact of causal trait theory generation on pattern projection. When memory (as well as the higher-order interactions) was added to the model, an accurate memory for trait-based information about the target did not enhance the chance that participants would pattern project from targets, B = 29.15, SE = 20.74, t(90.89) = 1.41, p > .16. Furthermore, the Patterning X Processing Task interaction, as well as the Patterning X Processing Task X Difficulty interaction, remained significant: t(849.28) = 1.94, p = .05, and t(2,401.83) = 3.27, p = .001, respectively. Instead, the (successful) generation of a causal trait theory *causes* pattern projection: In support of the quantity hypothesis, participants prompted to generate causal trait theories about others began to pattern project from them.

2.9 Study 4

We have argued that causal trait theories underlie pattern projection because an explanation that accounts for trait co-occurrence within a single person can be easily exported to be a more general theory of human personality—i.e., an implicit personality theory. But Study 2 found that the greater quantity of theories to explain the self vs. someone else did not fully explain the reason why pattern projection is egocentric. For our final study, we differentiate two origins of causal trait theories, a distinction that will ultimately help explain why pattern projection is egocentric.

We suggest that causal trait theories are sometimes prompted upon observing a person's behavior in a *single context* (e.g., "She is being very talkative but not very polite right now.") But in other cases, causal trait theories are created after reflecting on why different trait-relevant behaviors a person shows across *multiple contexts* co-exist (e.g., "He is very talkative, much as he was at last night's party, but also not that polite, like when he was curt with the waiter.") Single context theories are in the service of making sense of behavior one has observed, whereas multiple context theories are in the service of making sense of one's overall personality. We suggest that multiple context theories, as true theories about personality (instead of about why trait-relevant behaviors would co-occur as they do in a specific situation), are more likely to be exported to characterize one's general implicit personality theories. In our vantage point as an outside observer, we would seem to be more likely to seek to explain why others behave as they do in a given context (i.e., create single-context theories). But given the self is, by definition, with itself in every context through which it lives, we expected people would be more likely to create multiple-context theories about the self. In combination, this would help explain another reason why pattern projection is egocentric.

2.9.1 Pilot Study #1: Do Causal Trait Theories for the Self (vs. Another) Skew Toward Multiple-Context Theories?

We explained to 132 Americans on Mechanical Turk what a causal trait theory is, and the distinction between a single-context and multiple-context theory. Participants answered two questions, indicating whether the extent to which causal trait theories they have to describe themselves, or someone else (e.g., a roommate, a coworker) tend to be multiple-context (1) or single-context (9) theories. Participants reported having relatively more multiple-context theories (vs. single-context theories) that explained the self (M = 4.37, SD = 2.02) as opposed to someone else (M = 5.10, SD = 1.96), paired $t(131)$ = 3.50, p = .001, d = .30.

2.9.2 Pilot Study #2: Are Multiple-Context (vs. Single-Context) Theories More Common in Explaining the Self?

A second pilot study explored the converse question. We described to 193 undergraduates at the University of California, Berkeley, the same three constructs: causal trait theory, single-context theory, and multiple-context theory. Again, participants answered two questions, indicating whether the multiple-context theories and single-context theories they hold are more likely to describe the self (1) or their freshman-year roommate (9). In this study, the endpoints were counterbalanced, but responses were coded to match this form. Offering convergent support for our proposal, participants reported that their multiple-context theories were more likely to feature the self (M = 4.17, SD = 2.04) than were their single-context theories (M = 5.22, SD = 2.09), paired $t(191)$ = 4.47, p < .001, d = .32.

In combination, the two pilot studies support our proposal that causal trait theories for the self vs. another tend to differ in the breadth of their origin. If it were shown that causal trait theories that draw upon information from multiple contexts lead to more pattern projection than those that draw upon information from a single context, then this would demonstrate a second reason why pattern projection is egocentric.

To test the breadth hypothesis, participants in Study 4 received behavioral information about (and provided by) a yoked participant. This information described how the yoked participants displayed their typical standing on two traits in a single context, or described two distinct contexts in which they displayed their typical standing on each trait individually. Participants then attempted to generate a causal trait theory on the basis of this *single-context* or *multiple-context* information. Our primary prediction was that participants would be more likely to pattern project those theories generated on the basis of multiple contexts than on the basis of a single context.

2.9.3 Method

Participants and design. Two hundred five undergraduates at the University of California, Berkeley, participated in a study for which they received course credit or $15. Each participant was yoked to one of 81 participants who participated in an earlier session run for the purpose of generating materials for the present study.

Procedure. In the stimulus-construction sessions, the yoked participants first rated themselves on 12 traits from 1 (*not at all*) to 11 (*extremely*). Next, participants were asked to recall and write about 18 different behavioral episodes from their lives. For 12 of those behavioral episodes, participants were supposed to write about a time that they displayed their typical standing on one of the 12 traits they had rated earlier. Thus, they supplied twelve distinct single-context episodes. For the other 6 behavioral episodes, participants were supposed to recall a single context in which they displayed their typical level on two specified traits. Thus, for each of these six pair of traits, we had behavioral information describing their typical standing on these traits that came from a *single context* or that came from *multiple contexts*. These yoked participants were reminded before providing each behavioral episode that they should provide a detailed description with the knowledge that someone else would be reading what they wrote.

When the new crop of participants came to the lab for the main study, they first rated themselves on the same 12 traits, from 1 (*not at all*) to 11 (*extremely*). Next, they received a packet with behavioral information that had been provided by one of the participants from the original, stimulus-generating set. The packet included the previous participant's ratings of his or her own personality as well as exactly half of the behavioral episodes the participant had recalled. Participants received 3 of the *single-context* memories—each a description of a time when the previous participant had displayed his or her typical standing on two different traits in the same context. Participants learned about the yoked participant's standing on the other 6 traits through 6 multiple-context memories. In this way, participants always learned about the traits in pairs, but it was varied whether the memories came from a single episode or multiple episodes. We yoked at least two participants to each previous participant and counterbalanced for which trait pairs the single-context and the multiple-context memories were provided.

Participants began by reading all of the information in the packet—the ratings and the behavioral memories. Next, we had participants go back through the packet, but focus their attention on one pair of traits at a time. Just to reiterate, for three trait pairs participants learned about the target from two distinct episodes (one for each trait), whereas for the other half of trait pairs their attention was focused on one single episode that reflected both traits. After participants reviewed this information, they were asked to attempt to do their best to generate a causal trait theory to explain the trait co-occurrence. Participants were given 90 seconds to type a theory.

After completing each theory, participants reported, "How difficult was it to create the theory?", as well as their confidence in it, "How confident are you that the explanation is accurate?" Both were made on scales from 1 (not at all) to 7 (completely). The two items were negatively correlated (r = -.55), so we created a difference score to reflect *difficulty* with the theory generation task.

After trying to generate all 6 causal trait theories, participants then completed a series of conditional and marginal probability judgments from which their implicit personality theories could be extracted. More specifically, participants answered conditional probability questions of the form, "If all you knew about someone was that they were [trait X], how likely is it that that person would be [trait Y]?" In addition, participants made marginal probability judgments of the form "What percentage of people would you say are [trait X]?" We extracted implicit personality theories in the same way as in Study 3 (see also Critcher & Dunning, 2009).

2.9.4 Results

Although all participants received behavioral information about the same six pairs of traits, some participants learned about the previous participant's standing on the two traits through behavioral information from *multiple contexts,* whereas others received information from a *single context*. In order to test whether causal trait theories generated with behavioral information from multiple contexts (versus a single context) led to more pattern projection, we tested a random-slope, random-intercept multi-level model predicting participants' implicit personality theories. We defined three Level-1 variables nested within trait pair: *behavioral-information* (+1 = multiple contexts, -1 = single context), *yoked-difference* (the absolute value difference in trait ratings—for a particular trait pair—for the previous participant to whom the participant was yoked), and *self-difference* (to control for pattern projection from the self). Finally, we included a random effect of participant, which controlled for each participant's general tendency to see traits as more or less positively correlated.

Of key interest was the Yoked-difference X Behavioral-information interaction term. This could tell us whether the degree of pattern projection from the yoked participant depended on the nature of the behavioral information received from that yoked participant. As predicted, this interaction term was significant, B = 89.26, *SE* = 40.34, *t*(1,160.74) = 2.21, *p* = .03, semi-partial R^2 = .0042. When participants generated causal trait theories about the past participant on the basis of multiple-context information, participants pattern projected from the yoked participant, B = 202.89, *SE* = 89.49, *t*(6,861.51) = 2.27, *p* = .02, semi-partial R^2 = .0007. However, when participants developed causal trait theories from single-context information, they did not, B = 24.47, *SE* = 88.00, *t* < 1.

Thus, by prompting causal trait theories about others based on information that more reflects the origin of such theories about the self (i.e., behavioral information from multiple contexts instead of a single context), participants began to pattern project from those others. However, an alternate account is that people may simply have more difficulty developing causal trait theories when they can only draw on information from one context. Our data suggest this alternative account is unlikely, in that we found that participants actually experienced *more* difficulty trying to generate causal trait theories from multiple contexts (M = -0.13, SD = 2.82) than from a single context (M = -0.67, SD = 2.83), t(1,014.00) = 3.98, p < .001.

To more conclusively show that pattern projection was traceable to differences in the breadth (single-context or multiple-context) of the behavioral information on which causal trait theories were based, and not to the difficulty of generating a causal trait theory based on behavioral information from a single-context vs. multiple contexts, we would want to include the difficulty composite as well as higher-level interaction terms in our model. In this analysis, the crucial Yoked-difference X Behavioral-information interaction remained significant, B = -93.19, SE = 40.22, t(1,125.36) = 2.32, p = .02, semi-partial R^2 = .0047. Notably, the Yoked-Difference X Difficulty interaction was also significant, B = -97.37, SE = 41.96, t(858.75) = 2.32, p = .02, semi-partial R^2 = .0062, which replicated a finding observed in Study 3 that participants pattern projected less to the extent that they had difficulty generating the causal trait theory.

2.9.5 Discussion

The present study complemented Study 3 in identifying a second factor—beyond the mere creation of causal trait theories—that underlies pattern projection. Two pilot studies and the main study combined to support the breadth hypothesis—that causal trait theories that draw on information across contexts (as causal trait theories about the self are especially likely to do) are more likely to lead to pattern projection than causal trait theories that draw on information from a single context (such as when one seeks to make sense of another’s behavior in a single episode).

These findings are also important because they address an alternative hypothesis that perhaps it is not causal trait theories, but instead memories of specific instances when trait-relevant behaviors co-occurred, that underlies pattern projection. This alternative would have predicted *more* (not less) pattern projection in the single-context condition. Some readers may wonder how the present findings square with Study 3's results, given that Study 3 participants pattern projected from others without receiving specific behavioral information about them. Note our claim is not that causal trait theories can only be created when specific behavioral information is learned. Instead, we argue that causal trait theories are more likely to be exported (i.e., pattern projected) when they reflect theories of how traits relate in a single person, not when they reflect speculation about how behaviors may co-occur in a single context. In actual social perception, we usually do not learn about people through abstract trait labels, but instead by observing their behavior across one or more contexts (Asch, 1946).

2.10 General Discussion

People do not think of people's personalities in merely descriptive terms ("I am not very wordy; I am ambitious") but in explanatory terms ("I am not very wordy because I am ambitious: Chatterboxes annoy others and get left out from things"). The present paper identified such explanations as *causal trait theories*. People tend to have more causal trait theories for the self than they do for others. Furthermore, these theories are more likely to be attempts to explain properties of the self observed across different contexts, as opposed to theories that attempt to explain behavioral co-occurrence in a single context. To be sure, people can create causal trait theories about others, but those theories are less likely to draw upon behavioral information from multiple contexts, and instead seek to explain why a person is behaving in the various ways that he or she is in a single context.

Beyond documenting the existence and exploring properties of causal trait theories, the present studies showed how this construct assists in resolving a lingering mystery in the social cognition literature. Critcher and Dunning (2009) provided evidence of egocentric pattern projection—a qualitatively novel way in which the self influences social judgment—but conceded that it was unclear exactly why pattern projection arises. By introducing the notion of causal trait theories, the present studies identify the aspect of person perception that gives rise to pattern projection and specify how differences in the quantity and origin of self-knowledge vs. social knowledge explain why pattern projection is egocentric.

At its core, our account reasoned that causal trait theories—explanations for why two traits co-occur in a sample of one—might be generalized to become a theory of how two traits tend to relate in people in general. Consistent with this account, people pattern projected more (from the self or from an other) when people reported having a causal trait theory to explain why two traits co-occurred as they did in that person (Study 2). Because people reported a larger *quantity* of causal trait theories about the self than about someone else, this suggests one reason that pattern projection is egocentric. And indeed, when people were prompted to generate causal trait theories (and thus not simply learn trait information) about another person, participants began to pattern project from that person (Study 3).

But the difference in number of causal trait theories does not fully explain the egocentric nature of pattern projection: Study 2 found that causal trait theories about the self are more likely to be pattern projected than causal trait theories about an other. Two pilot studies found that people generate causal trait theories about the self vs. about others in different circumstances, meaning such theories have *origins* in different types of information. Causal trait theories about the self are more likely to be theories that try to explain why one has two traits that emerge as they do across a breadth of contexts. This is a theory about one's personality that can then be generalized as a more general theory of personality. Causal trait theories about others are relatively more likely to be theories about why people behaved as they did in a single context. These are theories of why traits co-exist in a single situation, and as such are less easily exported as general theories of personality. And indeed, when people received behavioral information about yoked participants from multiple contexts, instead of a single context, they began to pattern project from them (Study 4).

Several features of our data suggested that causal trait theories are either special, or operate above and beyond other factors in their facilitation of pattern projection. First, causal trait theories do not explain pattern projection merely because they serve as a marker of *semantic overlap* between two traits. By this alternative account, traits that share more similarity in meaning (e.g., bashful and reserved) are more likely to occur similarly in a person, to be seen as correlated in people's implicit personality theories, and to be explained by a causal trait theory vs. not. We have already discussed how this alternative would be hard to square with our correlational study (Study 2), but it certainly cannot account for our experimental ones in which participants showed more or less pattern projection from others depending on whether they were prompted to think about another in the way they tend to think about the self (Studies 3 and 4).

Second, pattern projection does not merely stem from memories or knowledge of the *episodic co-occurrence* of two traits—i.e., memories of when two traits occurred together. In the single-context condition in Study 4, participants learned how two traits co-occurred in a single episode. If it were episodic co-occurrence, and not causal trait theories, that underlay pattern projection, we would have seen more, not the predicted less, evidence of pattern projection in the single-context condition—i.e., when causal trait theories were informed by such co-occurrence.

Third, pattern projection does not stem merely from two traits being connected as part of an *indirect theory*, two traits that are indirectly connected in a self-narrative. Instead, pattern projection requires that people have a theory of why two traits are directly linked. As reported in Footnote 7, we replicated Study 2's finding that the presence of causal trait theories in the self predicts pattern projection, but traits that were explained merely by a third-variable cause (an explanation not of why the two traits are directly linked, but of why they both are products of the same third variable) were not pattern projected more.

Finally, note that our explanation for egocentric pattern projection embraces a *weak* form of egocentrism—an account that does not claim that the self's advantage in pattern projection is necessary, but an account that identifies where the self's greater influence comes from. First, the self has more causal trait theories to understand the self. But as Study 3 showed, by prompting people to think about others in terms of causal trait theories (as opposed to in a more in depth but piecemeal manner), pattern projection from that other increased. Second, causal trait theories about the self are more likely to lean on behavioral information from a greater breadth of contexts. And when people's causal trait theories about others begin to rely on information from multiple contexts, the causal trait theories they create to understand those others are more strongly pattern projected.

2.10.1 Why are there more causal trait theories to explain the self than to explain another?

But if this egocentrism can be overcome, why isn't it? Although the present paper shows that one reason pattern projection is egocentric is the larger number of causal trait theories to explain the self than someone else, we did not address why exactly it is that people engage in more theorizing for the self than for others. Below, we consider three factors that might explain this difference and assess the plausibility of each:

The self thinks about itself more. One intuitively appealing answer, but one that we find ultimately incomplete, is that the self is the object of its own thoughts more than are other people. In all of that egocentric thought, there would simply be more of a chance for people to elaborate causal narratives to understand the self. Although there is no doubt some truth in this statement, it seems unlikely to offer a complete explanation. For example, Critcher and Dunning (2009) gave false personality feedback on fictitious personality dimensions (V/Z dominance, front-/back-brainedness) about the self or about someone else. Even though participants had the same amount of (limited) time to consider this information about the self than about someone else, people still pattern projected this newly-learned information more when it was said to describe the self instead of someone else.

The self is motivated to explain the greater cross-situational variability it observes in itself vs. others. A second possibility, but one that we also ultimately reject, is that the nature of the self's own trait-based understanding more naturally lends itself to causal trait theory construction. The better we get to know a person, the more we observe inconsistencies in his or her behavior (Prentice, 1990). It should thus be relatively unsurprising that the self sees more cross-situational variability in its own behavior than in others' (Monson et al., 1980). At first glance, it might seem that such observed variability might facilitate the construction of causal trait theories. That is, understanding what cues are present (or absent) when one displays (or does not display) a trait may help to explain why the trait emerges. The problem with this intuition is that these causal antecedents are most likely going to be variable aspects of a situation instead of stable aspects of the person. And in fact, as Study 1c showed, people are more likely to develop causal trait theories to explain stable traits. Thus, the fact that people have more causal trait theories to explain the self, even as they observe more cross-situational variability in the self, speaks to just how impressive the egocentric nature of causal trait theorizing is. But also, research suggesting that people see less cross-situational variability for their own internal or covert traits (Goldberg, 1981) gives some hint as to which traits are most likely to be included in causal trait theories. Future research is needed to more fully explore for what types of traits causal trait theories are likely to emerge.

Causal trait theorizing satisfies self-understanding more than social understanding. A third possibility, which we see as the most plausible candidate of the three, is that people may be more motivated to construct theories of themselves due to the different functions of self-knowledge versus social knowledge. In particular, social knowledge may be amassed with an eye toward prediction, whereas self-knowledge may aim toward understanding. Knowing whether a friend is introverted is useful in determining whether or not to invite him to a raucous party. Trait labels provide simple, helpful summaries of others' preferences or dispositions. A causal trait theory, by contrast, is unlikely to yield the same predictive returns.

Self-knowledge is less in the service of behavioral prediction (given the self necessarily has less uncertainty about how it will behave), but is (at least in part) in the service of solving the basic epistemic goal of self-understanding or self-assessment (Festinger, 1954; Sedikides, 1993; Sedikides & Strube, 1997). Causal trait theories reflect a deeper form of insight than mere trait ascriptions. Furthermore, even when people do have deeper curiosities about others, they may find they need to generate fewer causal trait theories about another (as opposed to the self) before feeling that they have reached a level of deep understanding about that person. We look inward and see complex selves, with much hidden beneath the surface, whereas when we observe others' behavior, we feel it reflects a more complete picture of their personalities than it would of our own (Pronin, Kruger, Savitsky, & Ross, 2001). Thus, our desire for self-knowledge may push us not only to generate causal trait theories about the self to explain these intricacies beneath the surface, but to persevere in creating more theories before we reach the same level of epistemic satisfaction that we would were we creating the theories about another.

2.10.2 Comparing Causal Trait Theories to Life Narratives

Previous research has noted that people's sense of self moves beyond mere trait ascription toward fuller life narratives in an effort to see the self as a "coherent whole" and gain a full "understanding of ourselves and our goals and actions" (Baddeley & Singer, 2000, p. 200) in a way that permits people to see causal connections in their lives (Reese, Yan, Jack, & Hayne, 2010). The construction of a life narrative allows the self to be a storyteller (Bruner, 1990). Life narratives piece together important and meaningful episodes into a thread that offers a causally-coherent storyline of one's life and identity (e.g., Eagan & Thorne, 2010; King, Buston, & Geise, 2009). Furthermore, life narratives can be thought of as largely independent of traits—neither deriving from them (Bauer & McAdams, 2004), nor predicting life outcomes (e.g., subjective well being) in a way that is redundant with them (Bauer & McAdams, 2010).

The development of both causal trait theories and life narratives reflects a sense that the self is a more coherent and integrated entity than a list of descriptors might imply. Both involve extending beyond factual information to include interpretive information (see Pasupathi & Wainryb, 2010). Whereas life narratives attempt to draw semantic conclusions from *episodic* information about the self (McAdams & McLean, 2013), causal trait theories reflect deeper semantic engagement with prior *semantic* conclusions about the self (i.e., traits). And although life narratives at times seek to explain why a person has the traits he or she does today, such explanations lean on formative episodes in one's past instead of other traits in the self (Habermas & Block, 2000). In this sense, causal trait theories may reflect a deeper form of attribution—one that does not merely see "the person" as an attributional end in itself (Kelley, 1967; cf. Malle, Knobe, & Nelson, 2007), but that attempts to go one step deeper by explaining why "the person" is who he or she is.

Nonetheless, causal trait theories and life narratives may fulfill similar epistemic goals. Both permit people to create a coherent understanding of themselves instead of having merely disjointed person representations that lean on piecemeal descriptors. If causal trait theories and life narratives serve similar functions, it would be interesting to see whether those most likely to have well-elaborated narratives are more or less likely to have more causal trait theories. On the one hand, those with a greater orientation toward self-analysis and self-understanding may spend more time developing both. On the other hand, people may prefer to take one approach to self-knowledge or the other. That is, people may adopt a more artistic view by seeing the self as the protagonist in an on-going story (Bruner, 2002), or a more scientific view by seeing one's own traits and dispositions as mysteries that can be explained by other such traits and dispositions.

2.10.3 Interpersonal Trait Narratives, Interpersonal Pattern Projection

Causal trait theories need not refer only to intrapersonal trait dynamics. They could refer to interpersonal trait dynamics as well. People will include their conceptions of close others in the self (Aron, McLaughlin-Volpe, Mashek, Lewandowski, Wright, & Aron, 2004; Wright, Aron, & Tropp, 2002), and members of collectivist cultures may naturally have a more expansive view of the self (Markus & Kitayama, 1991). Thus, people's trait theories may expand to include features of others as well. Such theories may include explanations for why traits in the self relate to, influence, or are influenced by traits in close others.

Interpersonal trait theories may then produce pattern projection at the dyadic level. For example, romantic couples may co-construct theories for why a trait in one partner has given rise to a trait in the other (see Fivush, Bohanek, & Marin, 2010, for discussion of life narrative co-construction). The couple may then generalize these theories and use them as bases for expectations about new couples they meet. In this way, interpersonal pattern projection will look similar to standard pattern projection, except the two component traits reside in separate people instead of in the same person. But interpersonal pattern projection would also permit projection of the same traits (e.g., one partner's high neuroticism paired with the other partner's low neuroticism). These possibilities await test by future research.

2.10.4 Conclusion

The present research identified a new way we think about a person (i.e., the causal trait theory), used this construct to explain the emergence of pattern projection, and then explained why pattern projection emerges egocentrically. We suspect that the value of causal trait theories need not end with pattern projection, but may persist as an interesting construct to examine further in its own right. Much as psychologists spent decades understanding how people make attributions about the causes of behavior, there is clearly much to understand about how people make attributions about the origins of personality. Future research should look not only to better understand when and why these theories are formed, but also to identify additional outcomes that such theories predict.

2.11 References

Adler, J. M., & McAdams, D. P. (2007). Time, culture, and stories of the self. *Psychological Inquiry, 18*, 97-128.

Ahn., W., Marsh, J. K., Luhmann, C., & Lee, K. (2002). Effect of theory-based feature correlations on typicality judgments. *Memory & Cognition, 30*, 107-118.

Anderson, C. A., & Sedikides, C. (1991). Thinking about people: Contributions of a typological alternative to associationistic and dimensional models of person perception. *Journal of Personality and Social Psychology, 60*, 203-217.

Aron, A., McLaughlin-Volpe, T., Mashek, D., Lewandowski, G., Wright, S. C., & Aron, E. N. (2004). Including close others in the self. *European Review of Social Psychology, 15*, 101-132.

Asch, S. E., (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 303-314.

Asch, SE., & Zukier, H. (1984). Thinking about persons. *Journal of Personality and Social Psychology, 46*, 1230-1240.

Asch, S. B. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,*258-290.

Baddeley, J., & Singer, J. A. (2010). A loss in the family: Silence, memory, and narrative identity after bereavement. *Memory, 18,* 198-207.

Bauer, J. J., & McAdams, D. P. (2010). Eudaimonic growth: Narrative growth goals predict increases in ego development and subjective well-being 3 years later. *Developmental*

Psychology, 46, 761-772.

Beauregard, K. S, & Dunning, D. (1998). Turning up the contrast: Self-enhancement motives prompt egocentric contrast effects in social judgments. *Journal of Personality and Social Psychology 74*, 606-621.

Beer, A., & Watson, D. (2008). Asymmetry in judgments of personality: Others are less differentiated than the self. *Journal of Personality, 76*, 535-559.

Borkenau, P., & Liebler, A. (1994). The factor structure of trait ratings depends on the extent of information available to the judges. *European Review of Applied Psychology, 44*, 3–7.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Bruner, J. (2002). *Making stories: Law, literature, life*. New York: Farrar, Straus and Giroux.

Chapman, L. J., & Chapman, J.P. (1967). Genesis of popular but erroneous diagnostic observations. *Journal of Abnormal Psychology, 72,* 193-204.

Chater, N., & Oaksford, M. (2005). Mental mechanisms: Speculations on human causal learning and reasoning. In K. Fiedler, & P. Juslin (Eds.), Information sampling and adaptive cognition. London: Cambridge University Press.

Critcher, C. R., & Dunning, D. (2009). Egocentric pattern projection: How implicit personality theories recapitulate the geography of the self. *Journal of Personality and Social Psychology, 97*, 1-16.

Dawes, R. M. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology*, 25, 1-17.

Dunkel, C. S., & Anthis, K. S. (2001). The role of possible selves in identity formation: A short-term longitudinal study. *Journal of Adolescence, 24,* 765–776.

Dunning, D., & Cohen, G. L. (1992). Egocentric definitions of traits and abilities in social judgment. *Journal of Personality and Social Psychology, 63,* 341–355.

Dunning, D., & Hayes, A. (1996). Evidence for egocentric comparison in social judgment. *Journal of Personality and Social Psychology, 71,* 213–229.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082–1090.

Eagan, J., & Thorne, A. (2010). Life stories of troubled youth: Meanings for a mentor and a scholarly stranger. In K. C. McLean, & M. Pasupathi (Eds.), *Narrative development in adolescence* (pp. 113-129). New York: Springer.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87,* 327-339.

Epley, N., Morewedge, C., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social psychology, 40,* 760-768.

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

Fivush, R., Bohanek, J. G., & Marin, K. (2010). Patterns of family narrative co-construction in relation to adolescent identity and well-being. In K. C. McLean, & M. Pasupathi (Eds.), *Narrative development in adolescence* (pp. 45-63). Springer.

Goldings. (1954). On the avowal and projection of happiness. *Journal of Personality, 23,* 30–47.

Habermas, T., & Buck, S. (2000). Getting a life: The emergence of the life story in adolescence. *Psychological Bulletin, 126*, 748-769.

Hamilton, D.L., Katz, L.B., & Leirer, V.O. (1980). Cognitive representation of personality impressions: Organizational processes in first impression formation. *Journal of Personality and Social Psychology, 39*, 1050-1063.

Hampson, S. E. (1998). When is an inconsistency not an inconsistency? Trait reconciliation in personality description and impression formation. *Journal of Personality and Social Psychology, 74*, 102–117.

Holmes, D. S. (1981). Existence of classical projection and the stress reducing function of attributive projection: A reply to Sherwood. *Psychological Bulletin, 90,* 460–466.

Judd, C. M., Kenny, D. A., & Krosnick, J. A. (1983). Judging the positions of political candidates: Models of assimilation and contrast. *Journal of Personality and Social Psychology, 44,* 952–963.

Katz, D., & Allport, F. (1931). *Students' attitudes*. Syracuse, NY: Craftsman Press.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57*, 227-254.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation*. Lincoln: University of Nebraska Press.

Kim, M.P., Rosenberg, S. (1980). Comparison of two structural models of implicit personality theory. *Journal of Personality and Social Psychology, 38*, 375–89.

King, L. A., Burton, C. M., & Geise, A. C. (2009). The good (gay) life: The search for signs of maturity in the narratives of gay adults. In P. L. Hammack, & B. J. Cohler (Eds.), *The story of sexual identity: Narrative perspectives on the gay and lesbian life course* (pp. 375-396). Oxford University Press.

Krueger, J., & Stanke, D. (2001). The role of self-reference and other referent knowledge in perceptions of group characteristics. *Personality and Social Psychology Bulletin, 27,* 876–888.

Kunda, Z., Miller, D. T., & Claire, T. (1990). Combining social concepts: The role of causal reasoning. *Cognitive Science, 14*, 551-577.

López, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development, 63*, 1070–1090.

Malle, B. F., Knobe, J., & Nelson, S. (2007). Actor-observer asymmetries in behavior explanations: New answers to an old question. *Journal of Personality and Social Psychology, 93*, 491–514.

Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224-253.

McAdams, D. P. (1985). *Power, intimacy, and the life story: Personological inquiries into identity*. New York: Guilford Press. McAdams, D. P. (2001). The psychology of life stories. *Review of General Psychology, 5*, 100–122.

McAdams, D. P. (1995). What do we know when we now a person? *Journal of Personality, 63*, 363-396.

McAdams, D. P. (2001). The psychology of life stories. *Review of General Psychology, 5*, 100-122.

McAdams, D. P., & McLean, K. C. (2013). Narrative identity. *Current Directions in Psychological Science, 22*, 233-238.

McNorgan, C. M., Kotack, R. A., Meehan, D. C., & McRae, K. (2007). Feature-feature causal relations and statistical co-occurrences in object concepts. *Memory & Cognition, 35*, 418-431.

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, *121,* 277-301.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in cognitive science: Vol. 2. Theory and applications* (pp. 23-45). Chichester, U.K.: Ellis Horwood.

Pals, J. L. (2006). Constructing the “springboard effect”: Causal connections, self-making, and growth within the life story. In D. P. McAdams, R. Josselson, and Lieblch (Eds.), *Identity and story: Creating self in narrative* (pp. 175–199). Washington, DC: American Psychological Association Press.

Park, B. (1986). A method for studying the development of impressions of real people. *Journal of Personality and Social Psychology, 51,*907-917.

Park, B., DeKray, M. L., & Kraus, S. (1994). Aggregating social behavior into person models: Perceiver-induced consistency. *Journal of Personality and Social Psychology*, *66*, 437-459.

Pasupathi, M., & Wainryb, C. (2010). On telling the whole story: Facts and interpretations in autobiographical memory narratives from childhood through midadolescence. *Developmental Psychology, 46*, 735-746.

Prentice, D. A. (1990). Familiarity and differences in self- and other-representations. *Journal of Personality and Social Psychology, 59*, 369-383.

Pronin, E., Kruger, J., Savitsky, K., & Ross, L. (2001). You don't know me, but I know you: The illusion of asymmetric insight. *Journal of Personality and Social Psychology, 81*, 639-656.

Reese, E., Yan, C., Jack, F. & Hayne, H. (2010). Emerging identities: Narrative and self from early childhood to early adolescence. In K. C. McLean, & M. Pasupathi (Eds.), *Narrative development in adolescence* (pp. 23-43). New York: Springer.

Risen, J. L., & Gilovich, T., Dunning, D. (2007). One-shot illusory correlations and stereotype formation. *Personality and Social Psychology Bulletin, 33, 1492- 1502.*

Rosenberg, S, (1976). New approaches to the analysis of personal constructs in person perception. In A. W Lanfield (Ed.), *Nebraska symposium on motivation* (pp. 179-242). Lincoln: University of Nebraska Press.

Ross, L., Greene, D., & House, P. (1977). The false consensus phenomenon: An attributional bias in self-perception and social perception processes. *Journal of Experimental Social Psychology, 13*, 279-301.

Sedikides, C. (1993)., *Assessment, enhancement, and verification determinants of the self-evaluation process*. *Journal of Personality and Social Psychology, 65,* 317–338.

Sedikides, C., & Anderson, C. A. (1994). Causal perceptions of intertrait relations: The glue that holds person types together. *Personality and Social Psychology Bulletin*, *20*, 294-302.

Sedikides, C., & Strube, M.J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. *Advances in Experimental Social Psychology, 29*, 209-269.

Wright, S. C., Aron, A., & Tropp, L. R. (2002). Including others (and groups) in the self: Self-expansion and intergroup relations. In J. P. Forgas & K. D. Williams (Eds.), *The social* *self: Cognitive, interpersonal and intergroup perspectives* (pp. 343-363). Philadelphia: Psychology Press.

2.12 Interim Discussion

People perceive the world differently depending on their personality, experience, and individual needs, desires and motivation. Chapter 2 demonstrated that people project their own self-understanding in others by creating causal trait theories. In considering a connection between behavior in different contexts (e.g., creativity in the classroom and extroversion in one's fraternity), people are developing a theory about a person. As Study 3 showed, causal trait theories are not situation-bound, but instead draw on information across different contexts, reflecting conclusions from a mini case study about personality dynamics. As a result, the stage is set for exporting that theory to one's more general social understanding.

However, people may also form impression of others from one situation-bound judgment in which personality inferences reflect conclusions only from that incident alone. For example, attribution research has identified cases in which external factors, such as a persons' overt behavior can be seen as sufficient informative of the person's underlying character or dispositions (Jones & Davis, 1965; Jones, Davis, & Gergen, 1961). In line with such research, Chapter 3 investigates conditions in which people draw personality inferences when presented with a person's moral dilemma judgment.

Interest in moral dilemmas transcends philosophy, psychology, and neuroscience into popular culture. Dilemmas appear in movies (Barish & Pakula, 1982), newspapers (Pinker, 2008, January 13), and popular books (Edmonds, 2013), and hundreds of thousands of people have participated in moral dilemma research (e.g., Cushman, Young, & Hauser, 2006). Hence, there have been many opportunities for people to discuss their dilemma judgments with others—yet, researchers have not considered the implications of such conversations. Do people infer personality traits from other people's moral dilemma judgments and use these inferences to inform social decision making? I propose that people infer the impact of affect and cognition on other peoples' moral dilemma judgments, and use these inferences to form impressions regarding others' personality traits. Rejecting harm makes one appear warm but less competent, whereas accepting outcome-maximizing harm makes one appear cold but more competent.

Chapter 3 – Judging Those Who Judge: Perceivers Infer the Roles of Affect and Cognition Underpinning Others' Moral Dilemma Responses

Abstract

Whereas considerable research examines antecedents of moral dilemma judgments where causing harm maximizes outcomes, this work examines social consequences: whether participants infer personality characteristics from others' dilemma judgments. We propose that people infer the roles of affective and cognitive processing underlying other peoples' moral dilemma judgments, and use this information to inform personality perceptions. In Studies 1 and 2, participants rated targets who rejected causing outcome-maximizing harm (consistent with deontology) as warmer but less competent than targets who accepted causing outcome-maximizing harm (consistent with utilitarianism). Studies 3a and 3b replicated this pattern and demonstrated that perceptions of affective processing mediated the effect on warmth, whereas perceptions of cognitive processing mediated the effect on competence. In Study 4 participants accurately predicted that affective decision-makers would reject harm, whereas cognitive decision-makers would accept harm. Furthermore, participants preferred targets who rejected causing harm for a social role prioritizing warmth (pediatrician), whereas they preferred targets who accepted causing harm for a social role prioritizing competence (hospital management, Study 5). Together, these results suggest that people infer the role of affective and cognitive processing underlying others' harm rejection and acceptance judgments, which inform personality inferences and decision-making.

Keywords: moral dilemmas, social perception, meta-perceptions, lay theories, affect and cognition

Judging Those Who Judge: Perceivers Infer the Roles of Affect and Cognition Underpinning Others' Moral Dilemma Responses

"Non-violence, which is the quality of the heart, cannot come by an appeal to the brain." – Mahatma Gandhi

"*The sign of an intelligent people is their ability to control their emotions by the application of reason.*" – Marya Mannes

Imagine a passenger jet has been hijacked by terrorists, and is now heading towards a densely populated urban center. Is it acceptable to shoot this plane down—including the innocent civilians on board—in order to prevent it from wreaking widespread carnage? In 2003 the German government decreed that doing so was acceptable. However, in 2006, the German courts overruled this decision, arguing that the German military is forbidden from harming civilians regardless of circumstances (Whitlock, 2006). Imagine a discussion where one person supported the government's position, and another supported the court's position. What impressions do these decisions convey about each person: Who is warmer, and who is more competent? Who should be selected to work with children, and who to run a large organization?

The hijacked airplane dilemma is one example of a class of conundrums where causing harm maximizes overall outcomes. Philosophers (Foot, 1967) and lay people disagree (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001) over whether causing harm to maximize outcomes is the appropriate course of action. According to the dual process model, resolving such dilemmas depends on two psychological processes: affective reactions to harm drive harm rejection—consistent with deontological ethical positions where the nature of an action defines its morality (Kant, 1785/1959). Conversely, cognitive deliberation regarding costs and benefits drives harm acceptance—consistent with utilitarian ethical positions where the outcome of an action defines its morality (Mill, 1861/1998). Hence, in the hijacked airplane dilemma, people disapprove of shooting down the airplane based on their emotional reactions to that gruesome thought (e.g., sympathy for the victim, horror at the thought of committing murder).[10] Conversely, people approve of shooting down the airplane based on an abstract cost-benefit analysis regarding the total lives saved in each case, thereby logically deducing that causing harm results in the lesser two evils. A great deal of research supports the dual-process model of moral judgment (e.g., Bartels, 2008; Conway & Gawronski, 2013; Greene et al., 2004; Nichols & Mallon, 2006; Suter & Hertwig, 2011; c.f. Mikhail, 2007). However, researchers have examined primarily the antecedents of such judgments—comparatively little is known regarding their consequences, including social consequences.

One consequence may be that people's dilemma judgments influence how others perceive them. Haidt (2001) argued that moral judgments are social in nature: they communicate important information about the speaker. Are listeners picking up on this information, and inferring psychological processes behind the speaker's moral judgments? People appear quite sensitive to psychological factors driving other kinds of moral decisions (e.g., Cushman, 2008; Pizarro & Tannenbaum, 2011; Weiner, 1985). Recent work suggests that perceivers are indeed drawing personality inferences from others' dilemma judgments (Everett, Pizarro, and Crocket, 2016; Kreps & Monin, 2015; Uhlmann, Zu, and Tannenbaum, 2013). However, the question remains as to whether lay people infer the *processing* behind others' judgments—do they surmise that emotions compel people to reject and logic motivates people to accept outcome-maximizing harm?

[10] Note that we are not endorsing the strong version of dual process theory, which proposes that affective reactions occur quickly by default, and cognitive evaluations occasionally over-ride default them; it seems incorrect (e.g., Baron, Gürçay, Moore, & Starcke, 2012). We are endorsing the softer version of the dual-process model which suggests that affective reactions to harm and cognitive evaluations of outcomes independently predict dilemma judgments, regardless of temporal order (Conway & Gawronski, 2013).

We propose that people infer how affect and cognition underpin others' moral dilemma judgments, and use this information to draw inferences about others' warmth and competence. Specifically, perceivers should rate targets who make characteristically[11] deontological judgments (i.e. *causing harm is inappropriate regardless of outcomes*) as relatively warm, because they appear to experience stronger tenderhearted affective reactions to the thought of harming someone (consistent with research linking harm rejection judgments to empathic concern, e.g., Conway & Gawronski, 2013). Conversely, perceivers should rate targets who make characteristically utilitarian judgments (i.e. *causing harm is appropriate when it maximizes overall outcomes*) as relatively more competent, because they appear to engage in more dispassionate, outcome-focused cognitive processing that weighs various outcomes and selects the most favorable ones (consistent with research linking harm acceptance dilemma judgments to individual differences in reasoning and deliberation, e.g., Bartels, 2008; Royzerman, Landy, & Leeman, 2014).[12] However, these perceptions should only pertain when causing harm maximizes outcomes, rather than when people accept non-outcome-maximizing harm.

[11] The term 'characteristically' must be used because the terms *deontology* and *utilitarianism* refer to a variety of related philosophical perspectives that may not always align with this classification. Nonetheless, most theorists agree that deontological positions typically entail avoiding causing harm and utilitarian positions typically entail accepting causing harm on dilemmas where causing harm maximizes outcomes (Foot, 1967, Greene et al., 2001), so we retain this terminology.

[12] If the dual-process model is correct, responses to classic moral dilemmas do not perfectly reflect the degree to which decision-makers experience affective reactions or engage in cognition in an absolute sense. If classic moral dilemmas place affect and cognition in conflict, and ultimately judges may only choose one option, then judgments reflect the *relative* strength of each process. For example, accepting harm that maximizes outcomes may occur either due to strong cognition coupled with strong but slightly weaker affect, or weak cognition coupled with weaker affect. Hence, a judgment to accept causing harm does not reveal whether the judge experienced strong or weak affect—only that cognition outweighed whatever degree of affect they experienced. Nor does such a judgment guarantee that the judge engaged in strong cognition—only that whatever cognition they engaged in outweighed their affective experience. Some people may experience both extensive affect and extensive cognitive processing, whereas others engage in little of either. In order to estimate each processes independently, it is necessary to use a technique such as process dissociation (see Conway & Gawronski, 2013). However, in the current work we are not interested in the actual processes underlying dilemma judgments so much as lay perceptions of these processes. To that end, lay people, like many researchers, equate harm avoidance judgments with strong affect and harm acceptance judgments with strong cognition. This inference is effective as a rough heuristic, so long as researchers recognize that it does not perfectly describe moral dilemma processing.

Moreover, we predict that inferences flexibly operate in the other direction as well: people are capable of predicting dilemma decisions based on information about target processing styles. Specifically, people should expect sensitive, affective targets to reject harm, but rational, logical targets to accept outcome-maximizing harm. Finally, we predict that these inferences will influence subsequent social decision-making. For example, people should select targets who reject harm for social roles prioritizing warm, but select targets who accept harm for social roles prioritizing competence. We tested these hypotheses across six studies.

3.1 Warmth and Competence: Fundamental Dimensions of Social Perception

Traditionally, researchers have argued that perceptions of personality (e.g., Wiggins, 1979) and behavior (e.g., Wojciszke, 1994) involve two fundamental dimensions: how *warm* and how *competent* the target is (Judd, James-Hawkins, Yzerbyt, & Kashima, 2005). Although researchers use somewhat different taxonomies to describe these dimensions (e.g., communion/agency, Bakan, 1956, sociable/intellectual, Rosenberg, Nelson & Vivekananthan, 1968, other-profitable/self-profitable, Peeters, 1983, and morality/competence Wojciszke, 1998), they all appear to cohere with core warmth and competence constructs (Imhoff, Woelke, Hanke, & Dotsch, 2013).

Classically, warmth perceptions are theorized to track how benevolent targets appear to be, whereas competence perceptions generally track how effective targets appear to be at reaching their goals (Fiske et al., 2006).[13] Importantly, people tend to link warmth-related constructs—such as empathy, emotional expressivity, emotionality, and popularity—with an affect-laden, intuitive thinking style characterized by heuristic processing and emotional reactivity (Epstein, Pacini, Denes-Raj, Heier, 1996; Shilo, Salton, Sharabi, 2002; Norris & Epstein, 2011).

[13] We thank an anonymous reviewer for this suggestion.

Hence, when determining whether someone is warm, people may consider how much that person appears to experience affective reactions to the thought of causing harm, such as sympathy and compassion for victims or outrage at contemplating becoming a murderer. If so, then people may infer that a target is warm when that target makes judgments consistent with such affective reactions to harm (i.e., rejecting causing harm regardless of outcomes).

Conversely, people generally link competence-related constructs—such as ego strength, creativity, academic achievement, and self-esteem—with rational, systematic, cognitive processing (Epstein et al., 1996; Pacini & Epstein, 1999). Hence, when determining whether someone is competent, people may consider how much that person appears to engage in cognitive operations such as abstract cost-benefit analyses that weight five lives against one life. Conversely, people may infer that a target is competent when that target makes judgments consistent with rational, cognitive processing (i.e., accepting outcome-maximizing harm). Studies have even found that individual differences in need for affect influence warmth perceptions, whereas individual difference in need for cognition influence competence perceptions (Aquino, Haddock, Maio, Wolf, & Alparone, 2016). If so, then perceptions of affective processing ought to mediate the effect harm rejection on warmth judgments, and perceptions of cognitive processing ought to mediate the effect of harm acceptance on competence judgments.

Warmth and competence perceptions pertain not only to individuals, but also to social groups: people perceive stereotypically benevolent groups as warm and stereotypically powerful groups as competent (Fiske, Xu, Cuddy, & Glick, 1999; Fiske, Cuddy, Glick, & Xu, 2002; Cuddy, Fiske, & Glick, 2007; Imhoff et al., 2013). Moreover, people often match personality with group membership: People prefer recruiting warm people to fulfill stereotypically warm roles, and competent people to fulfill stereotypically competent roles (Rudman & Glick, 1999). Accordingly, we examined whether people show similar selection effects depending on inferences based on target dilemma judgments—selecting targets who reject harm for roles that prioritize warmth, whereas targets who accept harm for roles that prioritize competence.

3.2 Inferring Moral Processing

There is some preliminary support for the contention that people draw personality and processing inferences from moral dilemma judgments. Uhlmann and colleagues (2013) found that targets who made characteristically utilitarian judgments were perceived as more pragmatic, but lower in empathy than people who made characteristically deontological judgments. Kreps and Monin (2014) found that decision-makers who espoused deontological arguments were regarded as moralizing more than decision-makers who espoused utilitarian arguments—the latter appeared primarily pragmatic. Furthermore, Tetlock (2002) found that participants used the length of time others took to make a moral dilemma judgment as a cue to infer personality:

Agents who made the utilitarian choice quickly were evaluated more negatively, as this rapid response suggested they found it easy to endorse murder. Finally, Everett and colleagues (2016) found that perceivers view people who reject causing harm on classic dilemmas as more moral and trustworthy (and offered them higher endowments in trust games) than people who accept causing harm to maximize outcomes.[14] Together, these findings suggest that perceivers draw reliable personality inferences from others' moral judgments. What remains unclear is perceptions of processing: whether perceivers naturally understand the link between sympathetic affective reactions harm-rejection, and logical outcome-focused processing and harm-acceptance judgments. If so, this finding would suggest that ordinary people essentially intuit the dual-process model of moral judgments (Greene et al., 2001).

[14] Unless the person was fated to die and begged for mercy, or both targets appear equally destined for harm.

Moreover, alternative hypotheses are plausible and must be ruled out. Harm-rejection dilemma judgments are associated with deontological philosophy, which prioritizes rationality (Kant, 1785/1959), whereas harm-acceptance judgments are associated with utilitarian philosophy (Mill, 1861/1998), which is historically viewed as prioritizing emotions and sentiments (e.g., happiness) rather than logic (Kagan, 1998). Hence, it is plausible that lay people view harm-rejection judgments as indicative of high competence (due to serious rational thinking), whereas harm-acceptance judgments as indicative of warmth (due to affective concerns with happiness)—after all, accepting harm in moral dilemmas saves the most lives. Alternatively, perceivers might view both targets who accept and reject harm as low in both warmth and competence, given that both decisions result in serious harm to someone. We rule out these alternatives in the studies that follow.

3.3 Overview and Hypotheses

Across six experiments, we investigated whether people draw inferences about others' moral processing, and use this information in personality assessment, when predicting dilemma decisions, and when forming meta-perceptions. First, we predicted that people would rate targets as warmer when those targets rejected causing harm, and rate targets as more competent when those targets accepted causing harm (Studies 1 & 2). Second, we measured participants' perceptions of the target's affective and cognitive processing in order to clarify the mechanisms behind this effect: we expected that perceptions of affective processing would mediate the effect of dilemma decision on warmth ratings, whereas perceptions of cognitive processing would mediate the effect of dilemma decision on competence ratings (Studies 3a & 3b). Third, we hypothesized that people would use these lay theories to predict social decision-making (Study 4): people would expect sensitive, affective targets to reject causing harm, and expect rational, logical targets to accept causing (outcome-maximizing) harm (Study 4). Finally, we predicted that people would select targets who reject harm for warm roles, but select targets who accept harm for competent roles (Study 5).

3.4 Study 1

In Study 1, we examined whether participants rate targets who reject causing outcome-maximizing harm as warmer but less competent than targets who accept such harm. Following Fiske and colleagues (2002), we measured perceptions of morality as well. Some theorists conceptualize morality and warmth interchangeably (Wojciszke, 1998) whereas others have distinguished between them (Goodwin et al., 2014; Leach, Ellemers, & Barreto, 2007). We were agnostic as to whether participants would distinguish between perceptions of warmth and morality in the current work.

3.4.1 Method

Participants and design. We obtained 100 American participants (77 males, 23 females, M_{age} = 31.98, *SD* = 11.68) via Amazon's Mechanical Turk (www.amazon.com, 2014), who received $0.25. We aimed to obtain ~50 people per cell in all between-subjects designs. Although we did not conduct a priori power analyses, this heuristic resulted in ~90% power to detect the key contrast of interest across all studies (see p-curve analysis below). Participants were randomly assigned to one of two conditions (target decision: harm inappropriate vs. appropriate).

Procedure. Participants viewed a photo of a university student named Brad, and learned that Brad ostensibly responded to three moral dilemmas. Participants read each dilemma, presented on individual screens, in a fixed random order. Each dilemma entailed performing a harmful action in order to achieve a particular outcome. Each dilemma screen also presented Brad's ostensible judgment of whether causing harm was appropriate or not appropriate. Brad's answer was always consistent across dilemmas.

Next, participants rated Brad's warmth, competence, and morality using items adapted from Fiske and colleagues (2002). Participants indicated how well four warmth traits (*warm, good-natured, tolerant, sincere*), five competence traits (*competent, confident, independent,* *competitive, intelligent*), and one moral trait (*moral*) described Brad on 7-point scales anchored at 1(*not at all*) and 7(*very much*). Item order was randomized for each participant. We averaged judgments into composites of warmth (α = .87), competence (α = .82), and morality.

Moral Dilemmas. Participants viewed three incongruent (i.e., high-conflict, Koenigs et al., 2007) moral dilemmas ostensibly answered by Brad: the *Vaccine Dilemma* where a doctor contemplates administering a vaccine that will kill some patients but save many others, the *Crying Baby Dilemma* where smothering a baby will save other townspeople (both employing the exact wording from Conway & Gawronski, 2013), and the *Drug Lord Dilemma* which read:

> *You own a restaurant in a small South American town where a gang of drug dealers operate. They bring violence to the streets; several people from the community have lost their lives as a result of gang activity. One day the gang leader demands you make him a delicious meal. You know he is highly allergic to peanuts. You bring out two dishes: one without peanuts and one with enough peanuts to kill him. You could give the peanut dish to the leader, which would kill him but reduce gang violence in the area. Is it appropriate to serve the peanut dish to the gang leader in order to reduce gang violence, even though this will kill him?*

Each dilemma ended with a sentence describing a harmful action that would produce a specific outcome. After reading each dilemma, participants learned that Brad indicted either *yes, harm is appropriate* (consistent with utilitarianism) or *no, harm is not appropriate* (consistent with deontology). As the specific dilemma question asked impacts responses (Tassy, Oullier, Mancini, & Wicker, 2013) and therefore possibly inferences, we opted to employ the wording from Conway and Gawronski (2013), which was adapted from the original wording by Greene and colleagues (2001).

3.4.2 Results and Discussion

We submitted ratings to a 2 (target decision: harm inappropriate vs. appropriate) × 2 (personality measure: warmth competence) repeated-measures ANOVA with the first factor between-subjects and the last factor within-subjects (see Figure 1). There was no main effect of decision, $F(1, 98) = .79$, $p < .376$, $\eta_p^2 = .01$, 95% CI [0.00, 0.08], and no main effect of personality measure, $F(1, 98) = 3.13$, $p = .080$, $\eta_p^2 = .03$, 95% CI [0.00, 0.12]. However, the two-way interaction between target decision and personality measure was significant, $F(1, 98) = 49.36$, $p < .001$, $\eta_p^2 = .34$, 95% CI [0.22, 0.50]. Post-hoc tests demonstrated that participants rated Brad as warmer when he rejected ($M = 5.45$, $SD = 1.09$), versus accepted causing harm ($M = 4.50$, $SD = 1.20$), $F(1, 98) = 16.73$, $p < .001$, $\eta^2 = .14$, 95% CI [0.04, 0.27]. Conversely, participants rated Brad as less competent when he rejected ($M = 4.86$, $SD = 1.20$) versus accepted causing harm ($M = 5.49$, $SD = 0.80$), $F(1, 98) = 9.86$, $p = .002$, $\eta^2 = .09$, 95% CI [0.01, 0.21]. Morality ratings were similar to warmth judgments: Participants rated Brad as more moral when he rejected ($M = 5.80$, $SD = 1.30$) versus accepted causing harm ($M = 4.70$, $SD = 1.60$), $t(98) = 3.74$, $p < .001$, $\eta^2 = .12$, 95% CI [0.04, 0.23], which makes sense as warmth and morality were highly correlated (see Table 1). However, morality ratings also correlated moderately with competence.

These results provide initial evidence for our argument. Participants rated targets who rejected causing harm as relatively warm, suggesting they inferred that affect drove this characteristically deontological decision. Conversely, participants rated targets who accepted causing harm (to maximize outcomes) as relatively competent, suggesting that participants inferred that cognition drove this characteristically utilitarian decision.

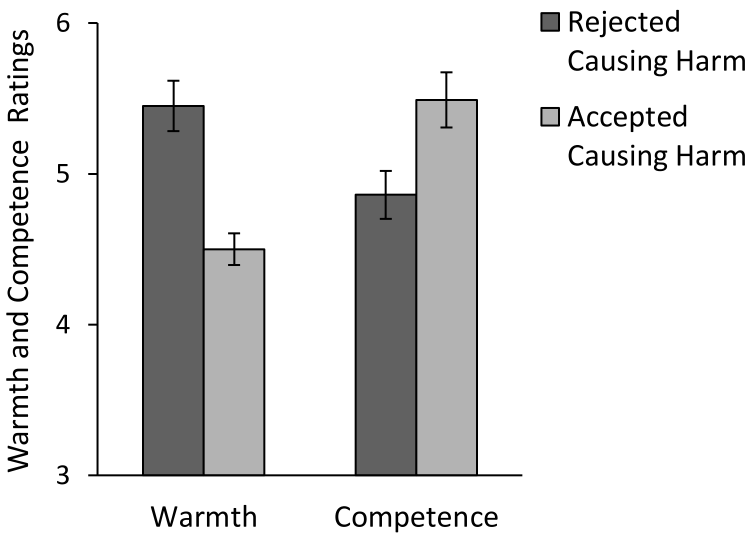


Figure 1. Target warmth and competence ratings when targets rejected or accepted causing harm to maximize outcomes, Study 1. Error bars reflect standard errors.

3.5 Study 2

Study 2 had two objectives: first, to examine whether the effects in Study 1 replicate, and to demonstrate one boundary condition. If, as we argue, perceivers draw inferences of affective and cognitive processing from target dilemma judgments, then modifying the particulars of the judgment ought to influence inferences. Classic dilemmas suggest a tension between emotion and reasoning because the (emotionally upsetting) decision to cause harm results in the (rationally objective) best overall outcome. Hence, people may infer that those who accept causing harm are competent, if not warm. Yet, consider findings that people high in psychopathy or with emotional deficits tend to accept causing harm on classic dilemmas (e.g., Bartels & Pizarro, 2011; Koenigs et al., 2007). This finding is likely driven by reduced affective reactions to harm, rather than increased concern with maximizing outcomes. Do lay people distinguish between bloodthirsty and genuinely utilitarian reasons for accepting causing harm?

One way to assess this possibility is by comparing inferences drawn from classic dilemma decisions to decisions where causing harm fails to maximize outcomes (though satisfies non-utilitarian motives, such as selfishness or vengeance). For example, if killing the baby no longer prevents the death of the other townsfolk, but instead prevents them from performing forced labor, killing the baby is no longer logically justified. In such cases, harm-accepting targets should no longer appear higher in competence than harm-rejecting targets. Moreover, they should appear especially low in warmth as such a decision reflects a pure lack of sympathy for the victim. In other words, if perceptions of competence reflect perceptions of outcome-focused logical reasoning, then only targets who accept causing outcome-maximizing harm should be rated as more competent than targets who reject causing harm or accept causing harm that fails to maximize outcomes. We examined these predictions in Study 2.

3.5.1 Method

Participants and design. We obtained 200 American participants (123 males, 77 females, M_{age} = 31.98, *SD* = 11.68) via Mechanical Turk, who received $0.25. Participants were randomly assigned to one of four conditions in a 2 (target decision: harm inappropriate vs. appropriate) × 2 (dilemma type: congruent vs. incongruent) between-subjects design, with warmth and competence ratings treated as within-subjects. Note this design subsumes a direct replication of Study 1 within the incongruent condition.

Procedure and materials. The procedure was identical to Study 1, except that participants read Brad's responses to either *incongruent* or *congruent* versions of the three moral dilemmas. Incongruent dilemmas correspond to classic, high-conflict dilemmas (Koenigs et al., 2007) where causing harm maximizes outcomes, and were identical to Study 1. Congruent dilemmas are worded identically to incongruent dilemmas, except that the positively consequences of harm are reduced, such that causing harm no longer maximizes outcomes. For example, in the congruent version of the Drug Lord dilemma, killing the drug lord will reduce mere car theft instead of gang violence. (i.e., participants viewed the exact same dilemma, replacing the words *drug dealers* with *car thieves*, *violence* with *car theft*, and *lost their lives* with *lost their cars*). Similarly, in the congruent Vaccine Dilemma, the deadly vaccine will cure only the common cold rather than a disease deadlier than the vaccine, and in the congruent version of the crying baby dilemma, killing the baby will merely prevent the soldiers from forcing the townsfolk to work instead of preventing the soldiers from killing them (for exact wording see Appendix A in Conway & Gawronski, 2013). After reading Brad's response to each dilemma, participants rated his warmth (α = .90), competence (α =.85), and morality via the same measures as in Study 1.

3.5.2 Results

Target warmth and competence. We submitted ratings to a 2 (target decision: harm inappropriate vs. appropriate) × 2 (dilemma type: congruent vs. incongruent) × 2 (personality measure: warmth competence) repeated-measures ANOVA with the first two factors between-subjects and the last factor within-subjects (see Figure 2). There was a main effect of decision, $F(1, 196) = 10.67$, $p < .001$, η_p^2 = .05, 95% CI [0.01, 0.12], no main effect of dilemma type, $F(1, 196) = 1.93$, $p = .167$, and a main effect of measure, $F(1, 196) = 11.50$, $p = .001$, $\eta_p^2 = .06$, 95% CI [0.01, 0.13]. These main effects were qualified by two-way interactions between target decision and personality measure, $F(1, 196) = 138.04$, $p < .001$, $\eta_p^2 = .42$, 95% CI [0.31, 0.50] and between dilemma type and personality measure, $F(1, 196) = 3.93$, $p = .049$, $\eta_p^2 = .02$, 95% CI [.01, .08], and between dilemma type and decision, $F(1, 196) = 26.98$, $p < .001$, $\eta_p^2 = .12$, 95% CI

[0.05, 0.20]. Moreover, the three-way interaction approached conventional levels of significance, $F(1, 196) = 3.53$, $p = .062$, $\eta_p^2 = .02$, 95% CI [0.00, 0.07]—close enough to cautiously consider tests of simple effects.

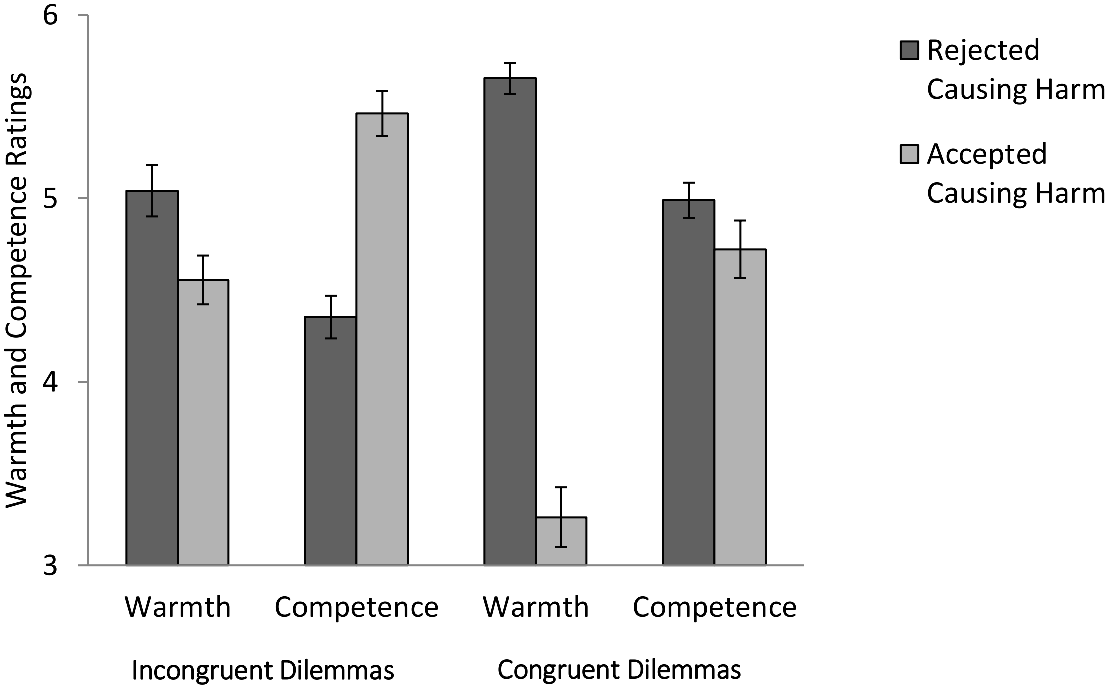


Figure 2. Target warmth and competence ratings when targets rejected or accepted causing harm when harm either maximized outcomes (incongruent dilemmas) or failed to maximize outcomes (congruent dilemmas), Study 2. Error bars reflect standard errors.

Post-hoc tests indicated that the pattern for incongruent dilemmas replicated Study 1: participants rated Brad as marginally warmer when he rejected ($M = 5.04$, $SD = 1.39$), rather than accepted ($M = 4.57$, $SD = 1.21$), causing harm, $F(1, 196) = 3.38$, $p = .068$, $\eta_p^2 = .02$, 95% CI [0.00, 0.06], whereas they rated him as less competent when he rejected ($M = 4.38$, $SD = 1.15$) versus accepted causing harm ($M = 5.44$, $SD = 1.11$), $F(1, 196) = 19.76$, $p < .001$, $\eta_p^2 = .10$, 95% CI [0.03, 0.17]. For congruent dilemmas, the pattern differed somewhat. Participants rated Brad as much warmer when he rejected (*M* = 5.62, *SD* = .85) versus accepted (*M* = 3.26, *SD* = 1.49) causing harm, *F*(1, 196) = 89.98, *p* < .001, η_p^2 = .32, 95% CI [0.21, 0.41], whereas they rated him as similarly competent whether he rejected (*M* = 4.97, *SD* = 1.05) or accepted causing harm (*M* = 4.72, *SD* = 1.41), *F*(1, 196) = 1.14, *p* = .29, η_p^2 =.01, 95% CI [0.00,0.04].

More importantly, simple-effects tests indicated that participants distinguished between Brad's reasons for causing harm: participants rated Brad as both warmer, *F*(1, 196) = 5.98, *p* = .015, η_p^2 = .03, 95% CI [0.00, 0.09], and more competent, *F*(1, 196) = 7.17, *p* = .008, η_p^2 = .04, 95% CI [0.00, 0.10], when he rejected causing harm on congruent, compared to incongruent, dilemmas. Conversely, participants rated Brad as both less warm, *F*(1, 196) = 24.45, *p* < .001, η_p^2 = .11, 95% CI [0.04, 0.20], and less competent, *F*(1, 196) = 8.23, *p* = .005, η_p^2 = .04, 95% CI [0.00, 0.11], when he accepted causing harm on congruent versus incongruent dilemmas.

Target morality. Finally, we also submitted target morality ratings to a 2 (target decision: harm inappropriate vs. appropriate) × 2 (dilemma type: congruent vs. incongruent) between-subjects ANOVA. Morality followed a pattern similar to warmth judgments. There was again no main effect of dilemma type, *F*(1, 196) = 2.55, *p* < .112, and a main effect of decision, *F*(1, 196) = 56.55, *p* < .001, η_p^2 = .22, 95% CI [0.13, 0.32], but this was qualified by a significant interaction, *F*(1, 196) = 24.04, *p* < .001, η_p^2 = .11, 95% CI [0.04, 0.19]. Post hoc tests for incongruent dilemmas replicated Study 1: participants rated Brad marginally higher in morality when he rejected (M = 5.44, SD = 1.64) versus accepted causing harm (*M* = 4.87, *SD* = 1.70), *F*(1, 196) = 3.35, *p* = .069, η_p^2 = .02, 95% CI [0.00, 0.07]. For congruent dilemmas, this pattern was amplified: Brad was rated much higher in morality when he rejected (*M* = 6.16, *SD* = 1.10) rather than accepted causing harm (*M* = 3.45, *SD* = 1.71), *F*(1, 196) = 78.92, *p* < .001, η_p^2 = .28, 95% CI [0.19, 0.38]. Again, warmth, competence and morality were positively correlated (see Table 1).

Table 1

Correlations between perceptions of warmth, competence, and morality, and perceptions of affective and cognitive processing in all studies.

Trait	Warmth	Competence	Morality	Affective Processing
Study 1				
Competence	.48***			
Morality	.75***	.40***		
Study 2				
Competence	.48***			
Morality	.85***	.50***		
Study 3a				
Competence	.31**			
Morality	.77***	.32***		
Affective Processing	.48***	-.25***	.46***	
Cognitive Processing	-.27**	.47**	-.27**	-.70***
Study 3b				
Competence	-.16			
Morality	.66***	-.23		
Affective Processing	.55***	-.49***	.36***	
Cognitive Processing	-.59***	.57***	-.35***	-.76***
Study 5 Pretest	.23*			

Competence

Note: * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

3.5.3 Discussion

Study 2 replicated the pattern from Study 1 for incongruent dilemmas: participants rated Brad as warmer and more moral but less competent when he rejected, rather than accepted, causing outcome-maximizing harm (although some contrasts were marginal). The pattern for congruent dilemmas was different: participants still rated Brad as (substantially) warmer and more moral when he rejected rather than accepted causing harm, but they no longer afforded him higher competence ratings for accepting causing harm. More importantly, participants drew different personality inferences depending on Brad's reason for causing harm: when Brad caused harm that failed to maximize outcomes (suggesting bloodthirstiness), they rated Brad as both less warm and less competent than when he caused outcome-maximizing harm.

This pattern suggests that participants were sensitive to *reasons* targets had for endorsing otherwise identical actions—targets who endorsed causing harm to maximize outcomes were viewed as more competent than those who endorsed causing the exact same harm for other reasons (e.g., selfishness, sadism). Moreover, the latter were also viewed as particularly low on warmth, suggesting that perceivers inferred that their harm-acceptance reflected a lack of regard for others' wellbeing similar to psychopathy (Hare, 1980). These findings suggest lay people distinguish between bloodthirsty and genuinely utilitarian reasons for causing harm, and competence perceptions track the latter.

3.6 Study 3a

In Studies 1 and 2, participants rated targets as warmer but less competent when those targets rejected, rather than accepted harm. These findings suggest that participants inferred the role of affect underpinning harm rejection, and the role of cognition underpinning (outcome-maximizing) harm acceptance. However, the model's predictions are more precise than just warmth and competence modulation: they suggest warmth ratings are driven by perceptions of affective processing, whereas competence ratings are driven by perceptions of cognitive processing. Hence, perceptions of target affective and cognitive processing should mediate the effect of target dilemma judgment on warmth and competence ratings, respectively.

In Study 3a, we again presented participants with Brad, who ostensibly accepted or rejected causing harm in the crying baby dilemma. We again measured participants' perceptions of Brad's warmth and competence, but first we directly measured perceptions of Brad's affective and cognitive processing. We expected to replicate Studies 1 and 2 regarding the impact of Brad's dilemma judgments on warmth, competence and morality perceptions. More importantly, we expected that perceptions of Brad's affective processing would mediate this effect on warmth ratings, whereas perceptions of Brad's cognitive processing would mediate this effect on competence ratings. Such a pattern would provide enhanced support for the claim that participants infer the roles of cognition and affect underpinning moral dilemma judgments, and that these inferences influence perceptions of the targets' warmth and competence. We were agnostic regarding whether affective or cognitive processing would mediate the manipulation on morality ratings.

3.6.1 Method

Participants and design. We recruited 121 American participants (71 males, 50 females, M_{age} = 36.12, *SD* = 11.88) via Mechanical Turk, who received $0.25. Participants were randomly assigned to one of two conditions (target decision: harm inappropriate vs. appropriate).

Procedure. The procedure was identical to Study 1, except we used only a single dilemma: participants viewed a photo of Brad and learned that Brad either ostensibly accepted or rejected harming the baby to save the townspeople in the crying baby dilemma. Next, participants indicated their perception of how much Brad's decision was based on *feelings and emotions,* as well as on *logical reasoning,* on scales from 1 (*not at all*) to 7 (*very much*). Finally, participants rated Brad's warmth (α = .89), competence (α = .86) as in Studies 1 and 2.

3.6.2 Results

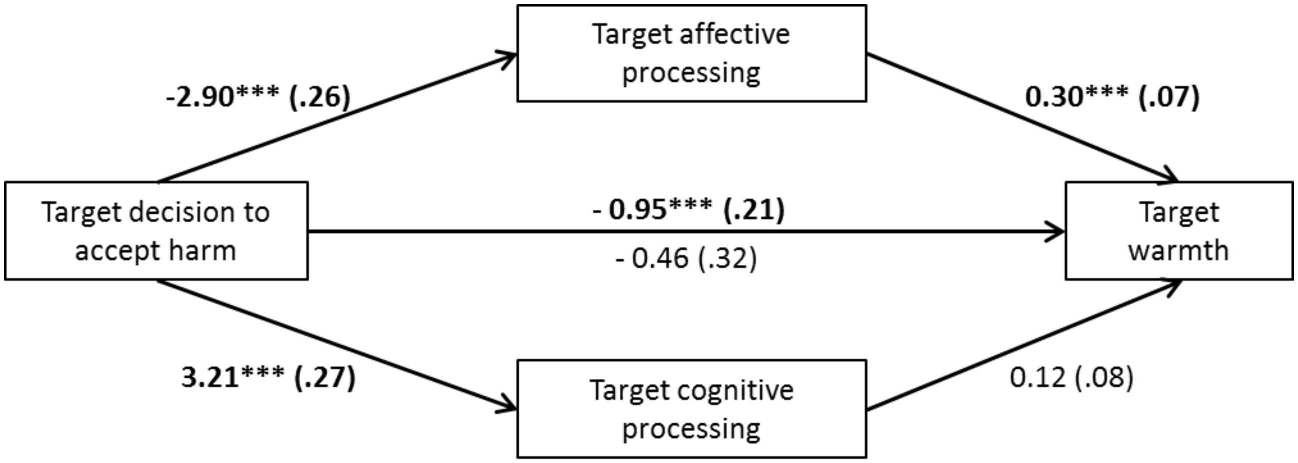
Target ratings. We submitted ratings to a 2 (target decision: harm inappropriate vs. appropriate) × 2 (personality measure: warmth competence) repeated-measures ANOVA with the first factor between-subjects and the last factor within-subjects (see Figure 1). There was no main effect of decision, $F(1, 119) = .15$, $p < .696$, $\eta_p^2 = .00$, 95% CI [0.00, 0.04], and no main effect of personality measure, $F(1, 119) = .48$, $p = .492$, $\eta_p^2 = .03$, 95% CI [0.00, 0.05]. However, the two-way interaction between target decision and personality measure was highly significant, $F(1, 119) = 80.45$, $p < .001$, $\eta_p^2 = .40$, 95% CI [0.27, 0.51]. Post-hoc tests replicated the finding that participants rated Brad as warmer when he rejected ($M = 5.38$, $SD = 1.12$) versus accepted causing harm ($M = 4.43$, $SD = 1.22$), $F(1, 119) = 18.06$, $p < .001$, $\eta_p^2 = .15$, 95% CI [0.05, 0.26]. We also replicated the finding that participants rated Brad as less competent when he rejected ($M = 4.43$, $SD = 1.23$) versus accepted causing harm ($M = 5.24$, $SD = .80$), $F(1, 119) = 18.23$, $p < .001$, $\eta_p^2 = .13$, 95% CI [0.08, 0.31]. Again, warmth, competence and morality were positively correlated (see Table 1).

Warmth mediation. To determine whether perceptions of affective or cognitive processing mediated the effect of Brad's dilemma decision on warmth ratings, we conducted a 10,000-iteration simultaneous mediation bootstrap analysis using the PROCESS macro according to the procedures recommended by Preacher and Hayes (2004, 2008). We then conducted identical analyses to determine whether perceptions of affective or cognitive processing mediated the effect of Brad's decision on both competence and morality. In the first step of each model, we regressed both affective and cognitive processing style on Brad's decision to accept or reject outcome-maximizing harm. As expected, Brad's decision to accept harm negatively predicted perceptions of affective processing, *B* = -2.90, *SE* = .26, *p* < .001, 95% CI [-3.42, -2.39], and positively predicted perceptions of cognitive processing, *B* = 3.21, *SE* = .27, *p* < .001, 95% CI [2.68, 3.75].

Next, we simultaneously regressed warmth ratings on Brad's decision and both mediators (see Figure 3a). As predicted, there was a significant indirect effect of decision on warmth through perceptions of affective processing, *B* = -.88, *SE* = .24, *p* = .001, 95% CI [-1.39, -0.45], but not cognitive processing, *B* = .40, *SE* = .28, *p* = .126, 95% CI [-0.20, 0.89]. A pairwise contrast of these effect sizes indicated that the indirect effect through affective processing was significantly larger than the indirect effect through cognitive processing, as the 95% confidence interval for the contrast excluded zero, *B* = -1.28, *SE* = .41, *p* = 0.01, 95% CI [-2.03, -0.39]. The direct effect of Brad's decision on warmth perceptions was not significant when both mediators were included in the model, *B* = -.46, *SE* = .32, *p* = 0.163, 95% CI [-1.09, 0.17]. These results indicate that participants thought Brad was warmer when he rejected versus accepted causing outcome-maximizing harm partly because they inferred he experienced stronger emotions. Although participants also inferred that Brad engaged in less cognition when he rejected versus accepted harm, perceptions of Brad's cognition had a weaker impact on warmth ratings than perceptions of affect. Rerunning this analysis controlling for competence or morality had little effect on results.

Competence mediation. We also simultaneously regressed competence on Brad's decision to accept harm and both mediators (see Figure 3a). As expected, the indirect effect through affective processing was not significant, *B* = -.31, *SE* = .19, *p* = 0.171, 95% CI [-0.69, 0.07], whereas the indirect effect through cognitive processing was, *B* = .85, *SE* = .30, *p* = .001, 95% CI [0.23, 1.43]. A pairwise contrast indicated that the indirect effect through affective processing was significantly smaller than the indirect effect through cognitive processing, *B* = -

1.15, *SE* = .40, 95% CI [-1.93, -0.29]. The direct effect of dilemma decision was not significant when the mediators were included in the model, *B* = .27, *SE* = .29, 95% CI [-0.30, 0.85]. Thus, participants thought Brad was more competent when he accepted versus rejected harming the baby because they inferred that he engaged in more cognitive processing. Although participants also inferred that Brad engaged in less affective processing when he accepted versus rejected harm, perceptions of Brad's affect had a weaker impact on competence ratings than perceptions of his cognition. Rerunning this analysis controlling for warmth or morality had little effect on results.



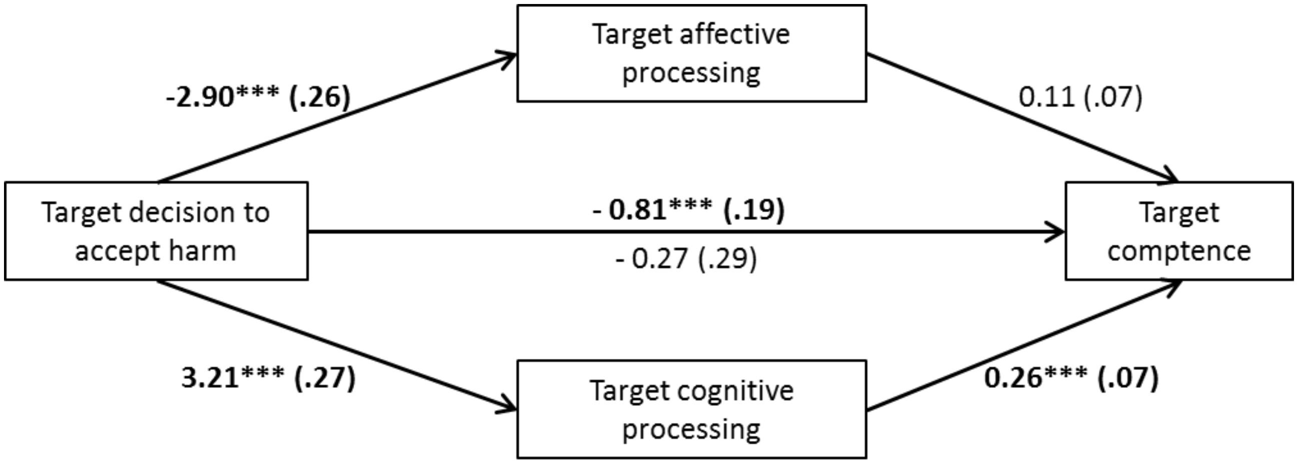


Figure 3a. Perceptions of the target's affective, but not cognitive, processing mediated the impact of that target's decision to reject causing harm on target warmth ratings, Study 3a. Perceptions of the target's cognitive, but not affective, processing mediated the impact of that target's decision to accept causing harm on target competence ratings, Study 3a. Unbracketed values reflect unstandardized coefficients; bracketed values reflect standard errors. * $p < .05$, ** $p < .01$, *** $p < .001$.

Morality mediation. Finally, we examined whether perceptions of affective or cognitive processing mediated the effect of Brad's decision on perceptions of his morality. There was a significant indirect effect through affective, $B = -.97$, $SE = .28$, 95% CI [-1.56, -0.48], but not cognitive processing, $B = .48$, $SE = .35$, 95% CI [-0.23, 1.15], and the indirect effect through affective processing was significantly larger than through cognitive processing, $B = -1.45$, $SE = .47$, 95% CI [-2.35, -0.51]. Meanwhile, the direct effect was not significant, $B = -.72$, $SE = .41$, 95% CI [-1.53, 0.09]. Taken at face value, this finding appears to indicate that people viewed Brad as more moral when he rejected harm because he appeared to engage more affective processing.

However, consider that warmth and morality were strongly correlated. Although controlling for morality had little effect on the warmth mediation, controlling for warmth reduced the indirect effect via affective processing to non-significance, $B = -.17$, $SE = .17$, 95% CI [-0.53, -0.13], while cognitive processing remained not significant, $B = .13$, $SE = .21$, 95% CI [-0.24, 0.58]. Controlling for competence had little effect. Therefore, perceptions of affective processing do not appear to directly mediate the effect of dilemma decision on morality. However, there remains the possibility that affective processing increases perceptions of warmth, which in turn increases perceptions of morality in a two-step mediation chain.[5]

Hence, we examined a mediation model where dilemma decision predicted affective processing, then warmth, then morality, controlling for logical processing. The indirect effect through this full pathway reached significance, *B* = -.47, *SE* = .18, 95% CI [-.91, -.20], whereas the indirect effect through affective processing alone did not, *B* = -.12, *SE* = .12, 95% CI [-.41, .08]; nor did the indirect effect through warmth alone, *B* = -.41, *SE* = .32, 95% CI [-1.04, .25]. Therefore, perceptions of affective processing indirectly mediated the impact of dilemma decision on morality ratings through increased perceptions of warmth, but neither perceptions of affective processing nor warmth directly mediated the impact of decision on morality. This pattern suggests that inferences of morality from rejecting harm on dilemmas are partly due to increased affective processing and therefore warmth, but warmth and morality are not redundant perceptions as perceptions of affect directly impact warmth but only indirectly impact morality.

3.6.3 Discussion

Overall, these findings enhance support for our argument. Not only did we replicate Studies 1 and 2 by finding that Brad's dilemma decision impacted perceptions of warmth, competence, and morality as expected—we also corroborated the precise hypothesis that perceptions of affective processing mediated this effect on warmth ratings, and perceptions of cognitive processing mediated this effect on competence ratings. This double-dissociation mediation pattern is especially remarkable considering that warmth and competence again correlated positively. Hence, these findings provide strong support for the argument that people infer the role of affective and cognitive processing underpinning dilemma judgments and used this information to inform personality perceptions.

Furthermore, we clarified the relation between warmth and morality. Again, these constructs were highly correlated and responded similarly to the manipulation, consistent with the possibility that they are interchangeable (Wojciszke, 1998). Yet, perceptions of affective processing mediated the impact of target dilemma judgment on perceptions of warmth, but only indirectly influenced perceptions of morality through perceptions of warmth. These findings suggest that warmth and morality are partly dissociable, though related, constructs (Goodwin et al., 2014). Next, we explored the robustness and generalizability of these findings by examining whether such a specific mediation pattern would replicate in a different sample, using dilemmas written in a different language, across both male and female targets.

3.7 Study 3b

Study 3b replicated Study 3a while enhancing generalizability in three ways. First, we employed a very different sample. Thus far, we had only tested American, English-speaking participants from Amazon's Mechanical Turk. Although initial evidence suggests that online samples are comparable to those obtained in the laboratory (Buhrmester, Kwang, & Gosling, 2011), we nonetheless wished to examine our predictions in a very different sample—specifically, German-speaking university students. Second, we employed all three dilemmas from Study 1, rather than the single dilemma used in Study 3a. Each dilemma was translated into German.

Third, we varied the gender of the target making moral decisions. Whereas previously participants rated only male targets, this study examined perceptions of both males and females. We introduced this manipulation because gender is a powerful predictor of moral dilemma judgments: women are much more likely than men to prefer the harm-rejecting response, likely due to stronger affective reactions to harm, whereas women and men are similar in their preference for the harm-accepting response, likely due to similar levels of cognitive processing (Friesdorf, Conway, & Gawronski, 2015). It is possible that participants' lay theories would reflect this trend by ascribing more affectivity, and hence more warmth, to female than male targets. However, we argue that warmth and competence perceptions should vary as a function of ascriptions of affective and cognitive processing. Gender differences emerge when averaging across many participants; when a particular man and woman make the same judgment for a particular dilemma, participants may reasonably infer that these targets experienced similar levels of affect and cognition. If so, then inferences regarding male and female moral processing should be similar. We compared these predictions in Study 3b.

We expected to replicate the mediation pattern from Study 3a: that perceptions of affective processing would mediate the impact of the target's dilemma judgments on warmth ratings, whereas perceptions of cognitive processing would mediate the impact of the target's dilemma judgments on competence ratings, and neither process would mediate the impact of judgment on morality ratings. We were agnostic as to whether this pattern would vary across gender. Replicating the mediation pattern from Study 3a in such a different sample, with different stimuli, and across target gender would increase confidence that the above-reported effects are robust to theoretically-irrelevant changes in stimulus materials or presentation context features.

3.7.1 Method

Participants and design. We recruited 120 students from the student cafeteria of a major German university (35 males, 85 females, M_{age} = 23.28, *SD* = 11.88) who participated in exchange for candy. Participants were randomly assigned to one of four conditions in a 2 (target gender: male vs. female) × 2 (target decision: harm inappropriate vs. appropriate) between-subjects design.

Procedure. The procedure was identical to Study 3, except that we presented participants with a picture of either 'Marc,' a male target, or 'Mara,' a female target (Emotionwisegroup, 2011). These targets displayed fairly neutral facial expressions and were rated as similar in attractiveness. After viewing the photograph, participants learned that Marc or Mara either consistently rejected or consistently accepted causing harm on the crying baby, drug lord, and vaccine dilemmas from Study 1. As in Study 3, participants indicated how much Marc or Mara's decision was based on both a) feelings and emotions, and b) logical reasoning, before rating Marc or Mara's warmth (α =.72), competence (α = .87), and morality. All measures and materials were presented in German.

3.7.2 Results

Target warmth and competence. In order to test the effect of target decision on warmth and competence we submitted ratings to a 2 (target decision: harm inappropriate vs. appropriate) × 2 (target gender: male vs. female) × 2 (personality measure: warmth vs. competence) repeated-measures ANOVA with the first two factors between-subjects and the last factor within-subjects.[15] There was a main effect of decision, $F(1, 114) = 17.40$, $p < .001$, $\eta_p^2 = .13$, 95% CI [0.04, 0.25], but no main effect of gender, $F(1, 114) = 2.64$, $p = .107$, $\eta_p^2 = .02$, 95% CI [0.00, 0.10] or measure, $F(1, 114) = .26$, $p = .613$, 95% CI [0.00, 0.05]. Neither the two-way interaction between gender and measure, $F(1, 114) = 2.43$, $p = .122$, $\eta_p^2 = .02$, 95% CI [0.00, 0.10], or decision and gender reached significance $F(1, 114) = .61$, $p = .437$, $\eta_p^2 = .00$, 95% CI [.00,.00], but replicating Studies 1 and 2, the interaction between decision and measure was significant, $F(1, 114) = 65.19$, $p < .001$, $\eta_p^2 = .37$, 95% CI [0.23, 0.48]. This was not qualified by a three way interactions, $F(1, 114) = 0.00$, $p = .994$, $\eta_p^2 = .00$, 95% CI [0.00, 0.00].

This pattern suggested that across both genders, participants rated targets who rejected ($M = 4.40$, $SD = 1.09$), rather than accepted causing harm ($M = 2.75$, $SD = .99$) as warmer, $F(1, 114) = 77.22$, $p < .001$, $\eta_p^2 = .40$, 95% CI [0.27, 0.51]. Conversely, participants rated targets who rejected ($M = 3.23$, $SD =1.05$), rather than accepted causing harm ($M = 3.91$, $SD = .95$) as less competent, $F(1, 114) = 13.82$, $p < .001$, $\eta_p^2 = .11$, 95% CI [0.02, 0.22]. In this case, warmth and competence were not significantly correlated, and even trended in opposite directions (see Table 1).

Target morality. Finally, we submitted morality ratings to a 2 (decision: harm inappropriate vs. appropriate) × 2 (target gender: male vs. female) between-subjects ANOVA.

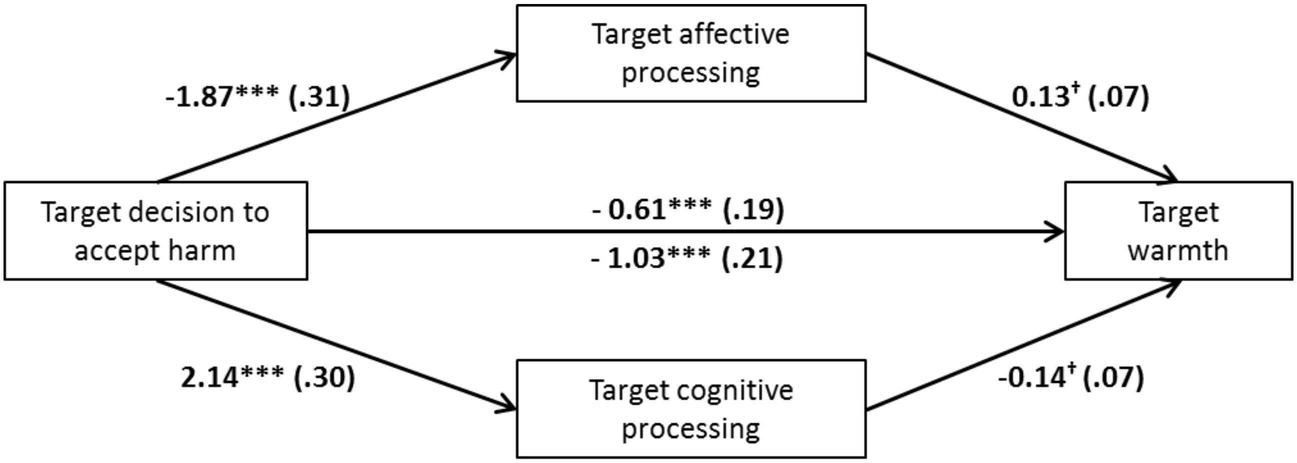
[15] Participants' gender did not interact with this finding, $F(2, 109) = .82$, $p = .367$, $\eta_p^2 = .01$.

We found the predicted main effect of decision: participants rated targets who rejected (*M* = 5.24, *SD* = 1.12), rather than accepted causing harm (*M* = 3.07, *SD* = 1.52), as more moral, *F*(1, 114) = 77.91, $p < .001$, η_p^2 = .41, 95% CI [0.27, 0.51]. There was no main effect of gender, *F*(1, 114) = 2.25, *p* = .137, η_p^2 = .00, 95% CI [0.00, 0.09], and the interaction was not significant, *F*(1, 114) = .91, *p* = .341, η_p^2 = .00, 95% CI [0.00, 0.09]. Warmth and morality again correlated positively, but morality did not significantly correlate with competence (see Table 1).

Warmth mediation. To determine whether perceptions of affective or cognitive processing mediated the effect of Brad's dilemma decision on warmth, competence, and morality ratings, we conducted analyses similar to those in Study 3a, with one key difference: As we also varied target gender in this study, we included this variable as a covariate (although doing so does not greatly affect results). In addition, five people were automatically dropped from this analysis due to minor amounts of missing data. Results are displayed in Figure 3b. In the first step of each model, we simultaneously regressed perceptions of the target's affective and cognitive processing on dilemma decision to accept or reject harm. As expected, the target's decision to accept harm negatively predicted perceptions of affective processing, *B* = -1.87, *SE* = .31, 95% CI [-2.48, -1.26], and positively predicted perceptions of cognitive processing, *B* = 2.14, *SE* = .30, 95% CI [-1.55, 2.74].

Next, we simultaneously regressed warmth on target decision and both mediators, controlling for target gender (see Figure 3b). We replicated the indirect effect through affective processing, *B* = -.25, *SE* = .15, 95% CI [-0.61, 0.01], although the effect was reduced to a marginal finding. Unexpectedly, we also found evidence for a marginal indirect effect through cognitive processing, *B* = -.30, *SE* = .18, 95% CI [-0.72, 0.01]. Although these indirect effects were in the same direction, they suggest slightly different interpretations: targets who rejected causing harm were perceived as warmer the more they appeared to engage in affective processing and the less they appeared to engage in cognitive processing. This time, these two effects were not significantly different from one another, *B* = .06, *SE* = .29, 95% CI [-0.53, 0.64]. The direct effect of decision on warmth also remained significant when both mediators were included in the model, *B* = -1.03, *SE* = .21, 95% CI [-1.45, -0.61]. Rerunning this analysis controlling for competence or morality had little effect on results.

Competence mediation. To examine the mediation of target decision on competence ratings, we again simultaneously regressed competence on Brad's decision to accept harm and both mediators, controlling for target gender and warmth (see Figure 3b). As predicted, and replicating Study 3a, the indirect effect through affective processing was not significant, *B* = .14, *SE* = .15, 95% CI [-0.13, 0.46], but the indirect effect through cognitive processing was, *B* = .53, *SE* = .19, 95% CI [0.21, 0.96]. However, the contrast between these effect sizes did not reach significance, *B* = -.40, *SE* = .31, 95% CI [-1.04, 0.19]. The direct effect of dilemma decision was also not significant when both mediators were included in the model, *B* = .06, *SE* = .20, 95% CI [-0.33, 0.45]. Rerunning this analysis controlling for warmth or morality had little effect on results.



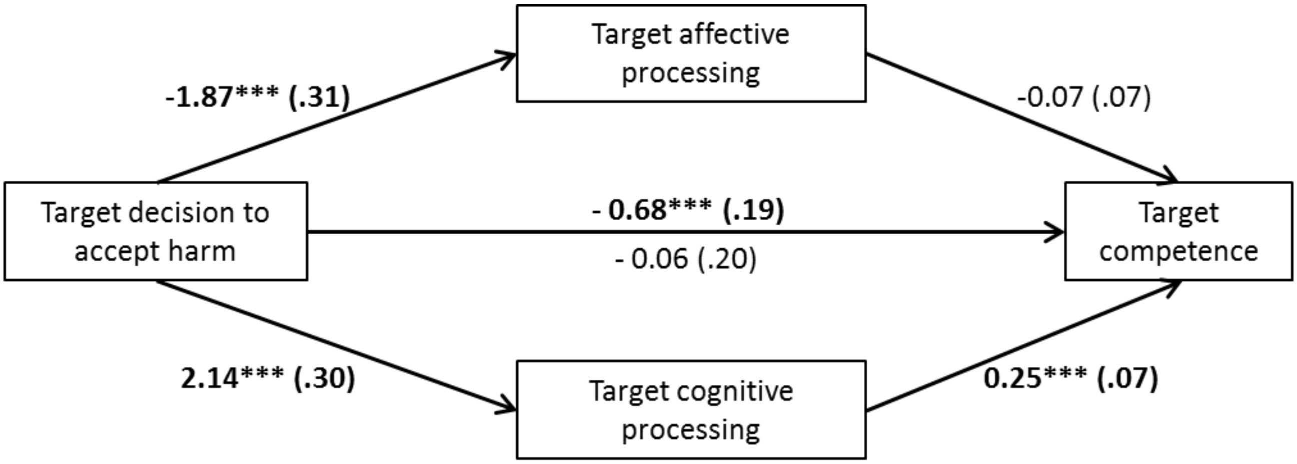


Figure 3b. Perceptions of both the target's affective and cognitive processing mediated the impact of that target's decision to reject causing harm on target warmth ratings, Study 3b. Perceptions of the target's cognitive, but not affective, processing fully mediated the impact of that target's decision to accept causing harm on target competence ratings, Study 3b. These findings remain similar while controlling for target gender. Unbracketed values reflect unstandardized coefficients; bracketed values reflect standard errors. † *p* = .06, * *p* < .05, ** *p* < .01, *** *p* < .001.

Morality mediation. Finally, we examined the mediation of target decision on morality ratings, controlling for warmth, competence, and target gender. Replicating Study 3a, this analysis revealed no significant indirect effect through affective processing, *B* = -.21, *SE* = .28, 95% CI [-0.76, 0.34], nor through cognitive processing, *B* = .24, *SE* = .32, 95% CI [-0.40, 0.90], and these effects were not significantly different from one another, *B* = -.44, *SE* = .57, 95% CI [-1.59, 0.67]. However, there remained a significant direct effect of Brad's decision on morality perceptions when both mediators were included in the model, *B* = -2.16, *SE* = .32, 95% CI [-2.78, -1.53]. This time, controlling for warmth or competence had little effect on results.

Again we assessed a chain mediation model from dilemma decision through affective processing to warmth to morality, controlling for logical processing. Unlike Study 3a, the indirect effect through the entire model did not reach significance, *B* = -0.04, *SE* = .04, 95% CI

[-.16, 0.01]. However, the indirect effect on morality through only warmth was significant, *B* = -0.64, *SE* = .26, 95% CI [-1.23, -0.23] and the indirect effect through affect alone was not, *B* = 0.01, *SE* = .06, 95% CI [-0.11, 0.16]. Although this pattern differed somewhat from Study 3a, it suggests a similar conclusion: although some of the impact of dilemma decisions on morality is carried through increased warmth ratings, perceptions of affective processing do not appear to contribute to this effect. Therefore, warmth and morality, although related construct, appear somewhat dissociable, with perceptions of affect mattering more for warmth than for morality ratings, and suggesting that other perceptions might influence morality ratings instead.

3.7.2 Discussion

These findings provide additional, albeit somewhat weaker, support for our argument. We replicated the pattern of target dilemma judgments on perceptions of target warmth, competence, and morality found in Studies 1, 2, and 3a. We obtained these effects in a very different sample—German-speaking university students in a cafeteria instead of English-speaking American online workers. Notably, findings held for both female and male targets. Gender did not impact results aside from higher overall ascriptions of warmth and competence to women than men. Unlike the previous studies, this time warmth and competence were not positively correlated. Perhaps this finding reflects cultural differences in notions of warmth and competence, the wording of translated materials, or simply statistical variation. Yet, despite this minor difference, we nonetheless replicated the effects of target dilemma decision on warmth, competence, and morality.

Moreover, we generally replicated the mediation pattern from Study 3a. As predicted, perceptions of affective processing (marginally) mediated the effect of decision on warmth ratings—however, this time perceptions of cognitive processing also (marginally) mediated the effect of decision on warmth ratings. This finding suggests that people rated targets who rejected causing harm warmer because such targets appeared to engage in both more affective and less cognitive processing. The competence mediation pattern from Study 3a replicated more clearly: people rated targets who accepted causing outcome-maximizing harm as more competent because these targets appeared to engage in more cognitive (but not affective) processing. However, this finding should be treated with caution as the contrast between these effect sizes did not reach significance. Finally, neither perceptions of affective nor cognitive processing mediated the effect of target decision on morality ratings, similar to the findings of Study 3a when controlling for warmth. Although in a chain mediation model warmth mediated some of the effect of decision on morality, perceptions of affective processing did not impact morality ratings the way they did warmth ratings, consistent with the possibility that somewhat different psychological processes underpin perceptions of warmth and morality (Goodwin et al., 2014). We return to this point in the general discussion.

3.8 Study 4

In Study 4 we tested whether participants infer that particularly affective targets should be more likely to reject causing harm, whereas particularly cognitive targets should be more likely to accept causing outcome-maximizing harm. To test this hypothesis, we again presented participants with Brad, but this time we gave participants only information about his decision-making style. When we described Brad as an affective decision-maker, we predicted that participants would expect him to choose *harm is not appropriate* more often. Conversely, when we described Brad as a cognitive decision-maker, we predicted that participants would expect Brad to choose *harm is appropriate* more often (as causing harm maximizes outcomes). In addition, we broadened the stimulus set beyond the dilemmas used previously by randomly presenting dilemmas from a standardized battery (Conway & Gawronski, 2013).

3.8.1 Method

Participants and design. We obtained 100 American participants (58 males, 42 females, M_{age} = 30.87, *SD* = 10.66) via Amazon's Mechanical Turk, who received $0.25.

Participants were randomly assigned to one of two conditions in a 2 (target decision-making style: affective vs. cognitive) × 2 (target decision likelihood: harm inappropriate vs. appropriate) repeated measures design, with the first factor between-subjects and the second factor within-subjects. Additional power afforded by the within-subjects component of the design enabled us to employ a lower per-condition sample size here than in the other studies.

Procedure. The procedure was a modification of Study 1. Participants viewed a photo of Brad, and a moral dilemma selected at random. Next, participants learned that Brad is either a rational or affective decision-maker. Specifically, they read: *Brad is a very rational [sensitive]*[16] *person who focuses on logic [his feelings] when making a decision*. *How likely is it that Brad selected each of the following decisions?* Participants then separately rated a) the likelihood that Brad decided causing outcome-maximizing harm is appropriate, and b) the likelihood that Brad decided causing such harm is inappropriate, on a 7-point scales anchored at 1 (*not at all likely*) and 7 (*very likely*). We used separate ratings to allow for the possibility that participants think a given decision-making style could increase motivation to *both* reject and accept causing harm. However, these measures were negatively correlated r = -.87, suggesting that participants generally viewed these decisions as opposites.

Instead of responding to the few dilemmas used previously, this time participants randomly viewed three of the ten incongruent (high-conflict) moral dilemmas from Conway and Gawronski (2013). Each dilemma entailed deciding whether to cause harm in order to maximize overall outcomes. The crying baby and vaccine dilemmas from Study 1 are examples of incongruent dilemmas from this set. Other examples include the *torture dilemma* (is it appropriate to torture a man in order to stop a bomb that will kill people?), and the *car accident*

[16] Note that instead of *emotional* we used the term *sensitive*, as not all emotions may lead to perceptions of warmth. For example, envisioning the target acting out of *anger* should probably not lead to a prediction of harm rejection.

dilemma (is it appropriate to run over a grandmother in order to avoid running over a mother and child?). We averaged participants' ratings across the three dilemmas they viewed.

3.8.2 Results and Discussion

We submitted likelihood ratings to a 2 (target decision-making style: affective vs. cognitive) × 2 (target decision: harm inappropriate vs. appropriate) repeated measures analysis with the first factor between-subjects and the second factor within-subjects (see Figure 4). There was no main effect of target decision-making style, $F(1, 98) = .49$, $p = .485$, $\eta_p^2 = .00$, 95% CI [0.00, 0.06], and no main effect of decision, $F(1, 98) = .40$, $p = .529$, $\eta_p^2 = .00$, 95% CI [0.05, 0.26]. However, there was a significant interaction, $F(1, 98) = 15.15$, $p < .001$, $\eta_p^2 = .13$, 95% CI [0.03, 0.26]. As predicted, post hoc tests indicated that participants in the affective decision-making condition thought it was more likely that Brad rejected ($M = 4.59$, $SD = 2.15$) than accepted causing harm ($M = 3.29$, $SD = 2.14$), $F(1, 98) = 5.43$, $p = .022$, $\eta_p^2 = .05$, 95% CI [0.00, 0.16]. Conversely, participants in the cognitive decision-making condition thought it was less likely that Brad rejected ($M = 3.12$, $SD = 1.99$) than accepted causing harm ($M = 4.92$, $SD = 1.97$), $F(1, 98) = 10.04$, $p = .002$, $\eta_p^2 = .09$, 95% CI [0.01, 0.21].

The results of Study 4 provided yet more support for our argument. As predicted, participants rated targets who engaged in affective processing as more likely to reject than accept causing harm, whereas they rated targets who engaged in cognitive processing as less likely to reject than accept causing harm. These findings suggest that people hold coherent lay theories regarding the connection between affective processing and harm rejection, and between cognitive processing and outcome-maximization. Moreover, we obtained this pattern across a set of 10 moral dilemmas rather than one or three. However, the question remains whether people's lay theories in this domain extend to more general decision-making. We examined this question in Study 5.

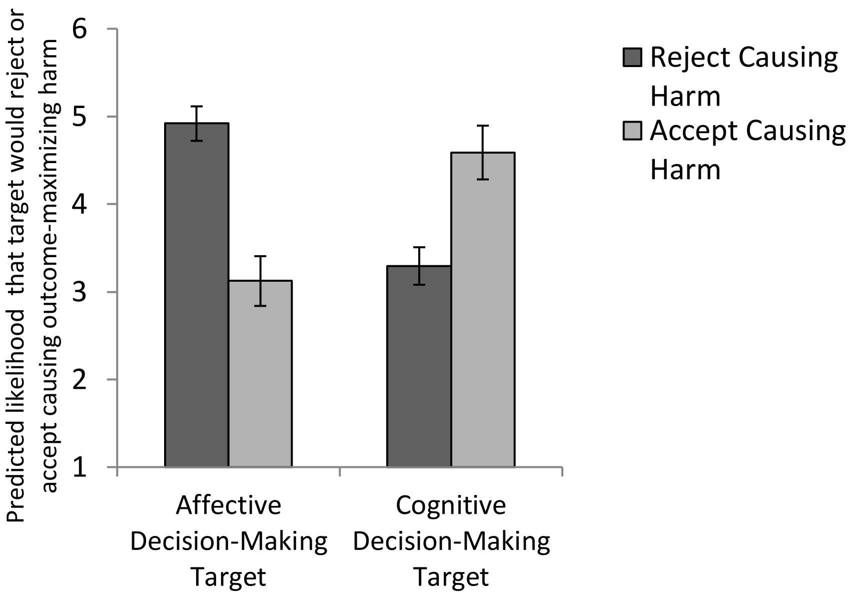


Figure 4. Participants' predicted likelihood that targets who prefer affective versus cognitive decision-making would accept or reject causing outcome-maximizing harm in various dilemmas, Study 4. Error bars reflect standard errors.

3.9 Study 5

Having demonstrated that people infer processing and personality traits based on others' moral dilemma judgments, we now examined whether these inferences impact behavior. Based on previous findings (Rudman & Glick, 1999; Cuddy & Fiske, 2002; Cuddy, Fiske, & Glick, 2004), we predicted that people would match targets to roles: they would select (warm) harm-rejecting targets for stereotypically warm social roles, and select (competent) harm-accepting targets for stereotypically competent social roles. To test this possibility, we introduced participants to Michael, a motorcyclist taking part in a motor cross race, who must decide whether to fatally shove a falling racer into a tree to prevent a deadly chain collision involving five other motorcyclists. Shoving entails killing the racer to save the lives of five other racers; not shoving entails avoiding killing the one racer (who will survive), but allowing the five other racers to die. In one condition Michael decided that it was appropriate to shove the racer; in the other condition Michael decided doing so was inappropriate.

Participants then rated Michael's suitability for the positions of pediatrician (where warmth is prioritized) and hospital manager (where competence is prioritized). If participants prefer Michael who rejects causing harm for the warmth-priority position and Michael who accepts causing outcome-maximizing harm for the competence-priority position, this would suggest that people use personality inferences based on moral dilemma judgments to inform social decision-making. However, one could of course argue that competence is important for both positions. Although we believe that this is true, we think that is mostly important that people value warmth *more* than competence for a pediatrician. That is, if both dimensions occur, warmth should carry more weight due to its primacy (Fiske et al.,2007). Hence, in order to ensure that people indeed prioritize warmth for doctors and competence for managers,we conducted a pretest.

3.9.1 Pretest

We recruited 120 American participants (68 males, 52 females, M_{age} = 33.87, *SD* = 11.78) from Amazon's Mechanical Turk to test whether participants prioritized warmth more for doctors than hospital directors, and competence more for hospital managers than doctors. We randomly assigned participants to imagine that either they were sick and seeking a doctor, or they were head of human resources and seeking a new hospital manager. We then asked participants: *Which traits should the doctor (future hospital manager) most importantly possess?* Finally, participants rated the importance of the warmth (α = .80) and competence (α = .51) items from Study 1 on scales from 1 (*not at all important*) to 7 (*very important*). Participants did not rate morality in this study.

We submitted these ratings to a 2 (role: pediatrician vs. hospital manager) × 2 (trait: warmth vs. competence) repeated measures analysis with the first factor between-subjects and the second factor within-subjects. There was no main effect of either social role, $F(1, 118) = .61$, $p = .438$, or trait, $F(1, 118) = .31$, $p = .574$, $\eta_p^2 = .00$, 95% CI [0.00, 0.00]. However, we found a significant interaction, $F(1, 118) = 28.20$, $p < .001$, $\eta_p^2 = .19$, 95% CI [0.08, 0.31]. As predicted, participants rated warmth as more important for a doctor ($M = 5.14$, $SD = .75$) than hospital manager ($M = 4.67$, $SD = .81$), $F(1, 118) = 10.61$, $p = .001$, $\eta_p^2 = .08$, 95% CI [0.01, 0.19], whereas they rated competence as less important for a doctor ($M = 4.80$, $SD = .57$) than hospital manager ($M = 5.10$, $SD = .56$), $F(1, 118) = 8.94$, $p = .003$, $\eta_p^2 = .07$, 95% CI [0.0, 0.17]. Moreover, participants seeking a doctor rated warmth ($M = 5.14$, $SD = .75$) as more important than competence ($M = 4.80$, $SD = .57$), $F(1,118) = 16.65$, $p < .001$, $\eta_p^2 = .12$, 95% CI [0.03, 0.24], whereas for participants seeking a hospital manager rated competence ($M = 5.10$, $SD = .56$) as more important than warmth ($M = 4.67$, $SD = .81$), $F(1,118) = 12.83$, $p = .001$, $\eta_p^2 = .10$, 95% CI [0.02, 0.21]. Note that warmth and competence again correlated positively (see Table 1), and all means were above the midpoints of the scales, suggesting that people value both warmth and competence for each position. Nonetheless, these findings indicate that people *prioritize* warmth for doctors and competence for hospital managers. Thus, if people use others' dilemma judgments to infer warmth and competence and thereby match people to social roles, then participants should select harm-rejecting Michael for the role of pediatrician, but harm-accepting Michael for the role of hospital manager.

3.9.2 Method

Participants and design. We obtained 107 German participants (15 males, 92 females, $M_{age} = 23.21$, $SD = 5.78$) from of a major German university, who received €2 or course credit. Participants were randomly assigned to one of two conditions in a 2 (target decision: harm acceptance vs. harm rejection) × 2 (social role: pediatrician vs. hospital manager) repeated measures design with the first factor between-subjects and the second factor within-subjects.

Procedure. Participants read the motorcyclist dilemma and learned about Michael's decision. Then, participants evaluated Michael's suitability (*How likely is it that Michael will be suitable for this position?*) for the pediatrician and hospital manager roles on separate scales from 1 (*not at all likely*) to 7 (*very likely*).

3.9.3 Results and Discussion

We submitted likelihood ratings to a 2 (target decision: harm inappropriate vs. appropriate) × 2 (role: pediatrician vs. hospital manager) repeated measures analysis with the first factor between-subjects and the second factor within-subjects (see Figure 5). There was no main effect of decision, $F(1, 105) = 1.87$, $p = .175$, $\eta_p^2 = .02$, [0.00, 0.09], but there was a theoretically uninteresting main effect of social role, $F(1, 105) = 16.25$, $p < .001$, $\eta_p^2 = .13$, 95% CI [0.04, 0.26]. More importantly, we found the predicted interaction, $F(1, 105) = 55.68$, $p < .001$, $\eta_p^2 = .35$, 95% CI [0.20, 0.47]. Post-hoc tests indicated that participants in the harm-rejecting condition thought that Michael was more suitable for the pediatrician ($M = 5.89$, $SD = 1.15$) than hospital manager position ($M = 3.61.50$, $SD = 1.86$), $F(1, 105) = 65.44$, $p < .001$, $\eta_p^2 = .38$, 95% CI [0.24, 0.50]. Conversely, participants in the harm-accepting condition thought that Michael was less suitable for the pediatrician ($M = 2.66$, $SD = 1.47$) than hospital manager position ($M = 4.57$, $SD = 1.69$), $F(1, 105) = 5.94$, $p < .016$, $\eta_p^2 = .05$, 95% CI [0.00, 0.15].

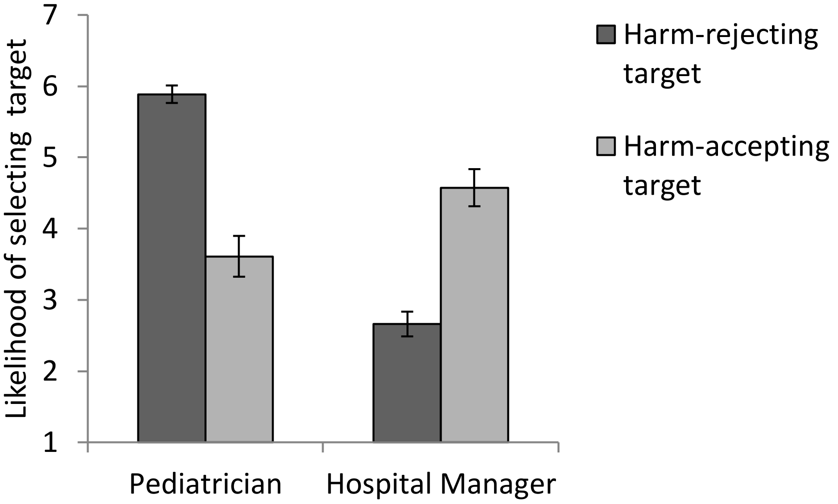


Figure 5. Perceived suitability of each target for the role of pediatrician or hospital manager when targets either rejected or accepted causing outcome-maximizing harm, Study 5. Error bars reflect standard errors.

These findings corroborate our argument that inferences of moral processing inform social decision-making. When people infer the roles of affective and cognitive processing underpinning others' moral dilemma judgments, they make predictions regarding social role fit. Specifically, people preferred to select a target who rejected causing harm (thereby suggesting they are a warm person) for a social role where warmth is prioritized. Conversely, people preferred to select a target who accepted causing harm (thereby suggesting they are a competent person) for a social role where competence is prioritized. Hence, participants appeared to match targets to social roles that fit their (inferred) personality traits, in line with previous work (e.g., Rudman & Glick, 1999). Note that in addition to the predicted interaction, there was also an overall main effect in favor of selecting the target that rejected causing harm. It may be that people generally prefer someone who makes a (warm) judgment, in line with research showing that warmth more important than competence (Fiske et al., 2006), or they generally trusted him more (Everett et al., 2016).

3.10 P-Curve Analysis

Finally, we conducted a p-curve analysis (Simonsohn, Nelson, & Simmons, 2014) using online p-curve software (www.p-curve.com/app3). This analysis assesses whether the statistical analyses in a body of work suggest the presence of a genuine phenomenon. The key prediction across Studies 1-3b and the Study 5 Pretest was that the impact of target dilemma decision on warmth is fully reversed on competence (for traditional, incongruent dilemmas); therefore, we included both of these simple effects for each study (Simonsohn et al., 2014). Similarly, for Studies 4 & 5 we again predicted full reversal as opposed to attenuation for the main dependent measure, so we again included these simple effects (see Table 2). This analysis indicated that across six studies, mean observed power was very high (95%). Analysis of right-hand skew (indicating many low p-values) indicates that the current studies contain evidential value, $Z = -10.16$, $p < .001$. Conversely, both the analysis of inadequate evidential value (indicating a lack of systematic findings), $Z = 7.01$, $p > .999$, and left-hand skew (indicating p-hacking), $Z = 10.16$, $p > .999$, failed to reach significance. Together these statistics indicate that the current studies were sufficiently well-powered.

Table 2

P-curve Analysis: Key Simple Effects Tests in Studies 1-5.

Trait/Condition	**Simple Test**
Study 1	
Warmth	$t(98) = 4.09$
Competence	$t(98) = -3.14$

Study 2	
Warmth	*F*(1, 196) = 3.38
Competence	*F*(1, 196) = -19.76
Study 3a	
Warmth	*t*(119) = 4.52
Competence	*t*(119) = -4.27
Study 3b	
Warmth	*F*(1, 114) = 77.22
Competence	*F*(1, 115) = -13.82
Study 4	
Affective Condition	*F*(1, 98) = 5.43
Cognitive Condition	*F*(1, 98) = -10.04
Study 5 Pretest	
Warmth	*F*(1, 118) = 10.61
Competence	*F*(1, 118) = -8.94
Study 5	
Harm Rejection Condition	*F*(1, 105) = 65.44
Harm Acceptance Condition	*F*(1, 105) = -5.94

3.11 General Discussion

Across six studies, we garnered evidence that people infer the affective and cognitive processing underpinning other' dilemma judgments. Participants rated targets who rejected causing outcome-maximizing harm on moral dilemmas as warmer but less competent than targets who accepted causing such harm (Studies 1-3). However, when targets accepted harm that failed to maximize outcomes, participants viewed them as colder and more immoral, but not more competent, than targets who rejected causing such harm (Study 2). Moreover, perceptions of affective processing mediated the effect of harm rejection on warmth, whereas perceptions of cognitive processing mediated the effect of harm acceptance on competence (Study 3a)—an effect that largely replicated in a different culture and across ratings of both male and female decision-makers (Study 3b). Moreover, people made the reverse inference: they expected targets who prefer affective processing to reject harm, whereas targets who prefer cognitive processing to accept harm to maximize outcomes (Study 4). Finally, people viewed harm-rejecting targets as more suitable for a social role prioritizing warmth, and harm-accepting targets as more suitable for a social role prioritizing competence (Study 5).

3.11.1 Implications

These findings add to a nascent but growing literature on perceptions of dilemma judgments by clarifying that perceivers not only infer qualities such as morality, pragmatism, and trustworthiness from others' moral dilemma judgments (Everett, et al., 2016; Kreps & Monin, 2015; Uhlmann et al., 2013)—but that perceivers essentially intuit the dual process model of moral judgments. People seem to be aware of the role that affective reactions to harm play in motivating harm rejection, and logical consideration of outcomes plays in harm-acceptance judgments (that maximize outcomes), and draw appropriate personality inferences. These findings may seem intuitive, but given historical associations between deontology and logic, and utilitarianism and emotion (Kagan, 1998), it was entirely plausible that lay people draw inferences along these lines instead. We also ruled out the possibility that people view all dilemma decision-makers as low on both competence and morality, given all of them produce sub-optimal outcomes.

The current findings may have far-reaching ramifications. Warmth and competence perceptions powerfully influence how people treat others—for example, warm others often induce pity, whereas competent others often induce respect (Fiske, 2002). Moral dilemma responses may lead to similar treatment: most people might prefer close interaction partners who reject harm, as they surmise these people are warm and good-natured. Indeed, Everett and colleagues (2016) demonstrated that most people—even many who themselves make the characteristically utilitarian judgments—prefer as social interaction partners people who selected characteristically deontological judgments. They speculated that deontological judgments serve as a signal of strong concern for others, hence trustworthiness, thereby attracting social partners (although whether characteristically deontological judgments serve as an *honest* signal of trustworthiness has yet to be determined). The current findings regarding perceptions of warmth and affective processing support that conclusion.

Yet, the current findings also add a caveat to Everett and colleagues' conclusions: people preferred others who made characteristically utilitarian judgments for the role of hospital manager, where raters prioritize competence over warmth. Hence, there may be social benefits to making both dilemma decisions, depending on what kind of social partner others are seeking. In most moral dilemma research, the characteristically utilitarian decision requires personally intervening (Gawronski et al., 2015), thereby personally shouldering the risk that one's intervention might be disastrous, whereas the 'characteristically deontological' decision entails stepping back and allowing fate to run its course (Gold, Colman, & Pulord, 2014). Hence, the characteristically utilitarian decision may provide evidence of leadership qualities (Lucas & Galinsky, 2015). Indeed, logical reasoning, complex problem solving skills, and emotional stability are important qualities for leaders (Friedman, Fleishman & Fletcher, 1992; Mumford, 2000), and the current work suggests people can demonstrate those qualities by accepting outcome-maximizing harm. Therefore, people may prefer targets who accept outcome-maximizing harm for leadership roles. Conversely, people may be averse to making the utilitarian choice when particularly worried about possible social sanctions (e.g., religious contexts, Szekely, Opre, & Miu, 2015), or when they need to build trust, such as online economic games (Everett et al., 2016). If so, this may explain why people make more characteristically 'utilitarian' judgments in the presence of ingroup members (Lucas & Livingston, 2014), as ingroup members may already trust decision-makers, freeing them to showcase other qualities.

Moreover, if people intuit the role of emotion and cognition in others' dilemma judgments, this raises the question of whether people are *aware* that others may judge them accordingly—and if so, whether they strategically shift dilemma judgments in order to create a desired impression. Indeed, some work suggests a role for motivated processing in moral judgments (Uhlmann, Pizarro, Tannenbaum, & Ditto, 2009; Liu & Ditto, 2014), and that dilemma judgments are sensitive to social influence (Kundu & Cummins, 2012). Such findings suggest that higher-order social cognitive processes may prospectively contribute to dilemma judgments beyond lower-order affective and cognitive processing. Future research should investigate this possibility.

We hasten to add that various factors may alter the demonstrated links between harm-rejection and emotion, and harm-acceptance and logic, if they alter inferences of the reasons why people act. We documented one such moderator in Study 2, where reducing the positive outcomes of harm reduced inferences of competence for targets who nonetheless still endorsed harm (as they appeared motivated by bloodthirsty rather than noble goals). Other manipulations that impact meaning of the dilemmas ought to have a similar effect on inferences. For example, greatly increasing the save-to-kill ratio reduces the need for cognitive processing to arrive at harm-acceptance decisions (Trémolière & Bonnefon, 2014); accordingly, we anticipate that participants would discount the competence of targets who accepting killing one person to save a million lives. Likewise, we expect that people would discount warmth inferences of people who refused such a trade-off, as their concern for saving one individual at the cost of millions suggests rigidity rather than genuine concern for well-being. Many other factors could potentially moderate the link between dilemma judgments and warmth/competence judgments if they implied different motivations for decision-makers. For example, the sex-ratio of the target and victims might invite ascriptions of romantic rivalry rather than moral processing (Trémolière, Kaminski, & Bonnefon, 2015), and the age (Kawai, Kubo, & Kubo-Kawai, 2014) or ethnicity (Uhlmann, Pizarro, Tannenbaum, & Ditto, 2009) of potential victims might invite ascriptions of prejudice rather than moral processing. Future work might profitably investigate some of these boundary conditions.

3.11.2 Moral Judgments and Judgments of Morality

The current findings are largely consistent with other recent findings regarding perceptions of dilemma decision-makers. However, one mystery emerged. Several studies have indicated that people view targets who reject causing harm as more moral than those who accept outcome-maximizing harm. Kreps and Monin (2014) found that people judged speakers who made purely deontic statements (“this is right”) as more moral than speakers who bolstered such statements with consequentialist justifications (“this is right *because*…”). Everett and colleagues (2016) demonstrated that people trust targets who reject harm more than those who accept harm. Ulhmann and colleagues (2013) found that people viewed targets who rejected harm as more moral than targets who accepted harm—an effect mediated through perceptions of empathic concern. One might anticipate that these findings obtain because such targets who reject harm appear to engage in more affective processing focused on the needs and well-being of others.

In the current work, we replicated the finding that people rated harm-rejecting targets as more moral than harm-accepting targets. However, perceptions of affective processing failed to directly mediate the impact of dilemma decision on mortality ratings. In chain mediation, perceptions of affective processing did not directly influence perceptions of morality—the only did so indirectly through perceptions of warmth in Study 3a. This suggests that although warmth may partially explain increased moral ratings of harm-rejections—consistent with Ulhmann and colleagues' (2013) mediation through empathic concern—nonetheless, morality ratings do not appear directly affected by perceptions of affect. Perhaps other inferred processes would better explain perceptions of morality. We suggest one possible candidate is perceptions of *moral rule adherence*. Nichols and Mallon (2006) argued that affective reactions to moral dilemmas depend upon appraisals that a moral rule has been violated. If lay perceptions accord with this model, then people should view targets who reject harm as both experiencing strong emotions and desiring to uphold moral rules. Perhaps the former primarily affects warmth, whereas the latter primarily affects morality, which would explain why the correlate highly yet nonetheless do not align perfectly. Future work should clarify these relations.

3.11.3 Limitations

In the current work, we relied upon traditional measures of warmth and competence (Fiske et al., 2006) which developed warmth and competence items via a 'top down,' theory-based fashion. This measurement model conceptualizes warmth as closely related to benevolence. Recent work assessing warmth and competence via 'bottom-up' data-driven approaches suggest that warmth might be more related to extraversion or sociability than benevolence, which better aligns with perceptions of morality (Goodwin et al., 2014). It could be that the warmth-as-benevolence items currently employed inflate the correlation between warmth and morality. It remains to be seen whether measures of sociability or extraversion evince the same patterns as warmth. However, re-analyses of all current data using only the terms 'warm' 'competent' and 'moral' provide a pattern of results similar to results using the full warmth and competence scales (see Supplementary Material). Future work should clarify the exact nature of warmth perceptions in moral judgment.

Like all moral dilemma research, the current work employed hypothetical moral dilemmas. Although such dilemmas appear to be useful tools to clarify the role of various psychological processes (Cushman & Greene, 2012), of course it remains unclear how well dilemma responses correspond to real-world decision-making. Nonetheless, the current work examines perceptions of others' dilemma responses, which may be similar for hypothetical and real dilemmas. Future work might clarify this by investigating perceptions of real-world decision-makers, such as the German judiciary that ruled on shooting down hijacked planes or Truman following his decision to use atomic weapons to end the Second World War.

Finally, the current studies contain two conflations common to the dilemma literature. First, participants in these studies appear to treat dilemma responses as reflecting pure processes (e.g., harm rejection judgments reflect strong affect and harm acceptance strong cognition), whereas, in fact the relation between responses and the processing underlying those responses is far more nuanced (Conway & Gawronski, 2013). However, as we are interested in lay perceptions of processes, rather than the processes themselves, this is a much weaker concern for the present work than for studies investigating these processes directly. The second conflation is that, like most of the literature, most dilemmas employed here conflate the characteristically utilitarian decision with taking action and the characteristically deontological decision with refusing to act (Gawronski et al., 2015). It is possible to consider scenarios where the reverse is true (e.g., intervening to save a person's life even though doing so places a larger number of others at risk). It seems plausible that people may draw a different pattern of inferences from decisions on such dilemmas.

3.11.4 Conclusion

The present work demonstrates that people infer the role of emotion and cognition underlying others' moral dilemma judgments. Recall the hijacked airplane dilemma from the introduction. The current findings suggest that the person who rejected shooting down the airplane seems warmer but less competent than the person who accepted shooting down the airplane, because they experienced stronger tender emotions and engaged in less cost-benefit reasoning. These findings open the door to the possibility that judgments in moral dilemma studies are driven not only by affective and cognitive processing, but also by strategic self-presentation. If so, then moral decision-making may be even more complicated than existing models suggest.

3.12 References

Abele, A. E., Cuddy, A. J. C., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology, 38*, 1063–1065. doi:10.1002/ejsp.574

Amazon (2011). https://requester.mturk.com/ Retrieved on February 18, 2014.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256.

Anderson, C., Ames, D. R., & Gosling, S. D. (2008). Punishing hubris: The perils of overestimating one's status in a group. *Personality and Social Psychology Bulletin, 34*, 90-101. doi:10.1177/0146167207307489

Aquino, A., Haddock, G., Maio, G. R., Wolf, L. J., & Alparone, F. R. (2016). The role of affective and cognitive individual differences in social perception.*Personality and Social Psychology Bulletin*, 0146167216643936.

Aristotle (1989/350BC). *Nichomachean ethics*. Oxford: Blackwell.

Bakan, D. (1956). *The duality of human existence: Isolation and communion in Western man*.

Chicago: Rand McNally.

Baron, J., & Ritov, I. (2009). Protected values and omission bias as deontological judgments. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.). *Moral judgment and* *decision making: The psychology of learning and motivation* (Vol. 50, pp. 133–167). SanDiego: Elsevier.

Baron, J., Gürçay, B., Moore, A. B., & Starcke, K. (2012). Use of a Rasch model to predict response times to utilitarian moral dilemmas. *Synthese,* 189, 107–117

Bartels, D. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition, 108*, 381–417. doi:10.1016/j.cognition.2008.03.001

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*, 154-161. doi:10.1016/j.cognition.2011.05.010

Beer, A., & Watson, D. (2008). Asymmetry in judgments of personality: Others are less differentiated than the self. *Journal of Personality, 76*, 535-559.

doi:10.1111/j.1467-6494.2008.00495.x

Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science, 5*, 187-202. doi:10.1177/1745691610362354

Bloom, P. (2011). Family, community, trolley problems, and the crisis in moral psychology. *The Yale Review, 99*, 26–43. doi:10.1111/j.1467-9736.2011.00701.x

Borkenau, P., & Liebler, A. (1994). The factor structure of trait ratings depends on the extent of information available to the judges. *European Review of Applied Psychology, 44*, 3-7.

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and sociobiology,13 171-195*. doi:10.1016/0162-3095(92)90032-Y

Boyd, R., & Richerson, P. J. (2005). Solving the puzzle of human cooperation. *Evolution and* *culture*, 105-132.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3-5. doi:10.1177/1745691610393980

Burnes, B. (2003). *Harry S. Truman: His life and times*. Kansas City, MO: Kansas City Star Books.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and social Psychology, 66*, 840-856. doi:10.1037/0022-3514.66.5.840

Christensen, J. F., Flexas, A., Calabrese, M., Gut, N. K., & Gomila, A. (2014). Moral judgment reloaded: a moral dilemma validation study. *Frontiers in Psychology, 5*, 1-18. doi:10.3389/fpsyg.2014.00607

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, *4,* 308-315.

Critcher, C. R., Dunning, D., & Rom, S. C. (2015). Causal trait theories: A new form of person knowledge that explains egocentric pattern projection. *Journal of personality and social psychology*, *108*, 400.

Connolly, P. (2009). *Ethics in action: a case-based approach*. John Wiley & Sons.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*, 216-235. doi:10.1037/a0031021

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92*, 631-648.

doi:10.1037/0022-3514.92.4.631

Cuddy, A. J., Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior*, *31*, 73-98.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*, 353-380.

Cushman, F., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience, 7*, 269-279.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*, 1082–1089.

doi:10.1111/j.1467-9280.2006.01834.x

Dawes, R. M. (1989). Statistical criteria for establishing a truly false consensus effect. Journal of Experimental Social Psychology, 25, 1–17. doi:10.1016/0022-1031(89)90036-X

De La Noy, K., Melniker, B. (Producers), & Nolan, C. (Director). (2008). *The Dark Knight* [Motion Picture]. United States: Warner Brothers.

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, *112*, 281-299.

doi:10.1017/CBO9780511808098.020

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological bulletin*, *139*, 477. doi:10.1037/a0029065

Eagly, A., H., & Karau, S. J. (2002). Role congrueity theory of prejudice towards female leaders. *Psychological Review, 109*, 573-598.

Edmonds, D. (2013). *Would you kill the fat man? The trolley problem and what your answer tells us about right and wrong*. Princeton, New Jersey: Princeton University Press.

Emotionwisegroup (Organization). (2011). Female target conveying happiness [Photograph]. Retrieved April 1st, 2014, from http://www.emotionwisegroup.org/wp-content/uploads/emotipedia/emotipedia/

Emotionwisegroup (Organization). (2011). Male target conveying happiness [Photograph]. Retrieved April 1st, 2014, from http://www.emotionwisegroup.org/wp-content/uploads/emotipedia/emotipedia/

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*, 327–339. doi: 10.1037/0022- 3514.87.3.327

Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General, 145*, 772.

Feinberg, M., Willer, R., Stellar, J., & Keltner, D. (2012). The virtues of gossip: reputational information sharing as prosocial behavior. *Journal of Personality and Social Psychology*, *102*, 1015. doi:10.1037/a0026650

Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological science*, *25,* 656-664. doi:10.1177/0956797613510184

Fischer, P., Greitemeyer, T., Pollozek, F., & Frey, D. (2006). The unresponsive bystander: Are bystanders more responsive in dangerous emergencies? *European Journal of Social Psychology*, *36*, 267-278. doi:10.1002/ejsp.297

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2006). Universal dimensions of social cognition: Warmth and Competence. *Trends in Cognitive Sciences, 11*, 77-83.

doi:10.1016/j.tics.2006.11.005

Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis) respecting versus (dis) liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, *55*, 473-489.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878-902.

doi:10.1037/0022-3514.82.6.878

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5,* 5-15. doi:10.1093/0199252866.003.0002

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in moral judgments: A process dissociation meta-analytic reanalysis. *Personality and Social Psychology Bulletin*.

Friedman, L., Fleishman, E. A., & Fletcher, J. M. (1992). Cognitive and interpersonal abilities related to the primary activities of R&D managers. *Journal of Engineering and Technology Management*, *9*, 211-242.

Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R., & Hütter, M. (2015). Moral dilemma judgments: Disentangling deontological inclinations, utilitarian inclinations, and general action tendencies. In J. P. Forgas, P. A. M. Van Lange, & L. Jussim (Eds.), *Social psychology of morality*. New York: Psychology Press.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*, 148-168. doi:10.1037/a0034726

Giangreco, D. M. (1997). Casualty projections for the U.S. invasion of Japan, 1945-1946: Planning and policy implications. *The Journal of Military History, 61*, 521-82.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology, 75*, 332-346. doi:10.1037/0022-3514.75.2.332

Gold, N., Colman, A. M., & Pulford, B. D. (2014). Cultural differences in responses to real-life and hypothetical trolley problems. Judgment and Decision Making, 9, 65-76.

Greene, J. D. (2003). From neural 'is' to moral 'ought': What are the moral implications of neuroscientific moral psychology? Nature Reviews, Neuroscience, 4, 847-850. doi:10.1038/nrn1224

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. Neuron, 44, 389-400. doi:10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. Science, 293, 2105-2108. doi:10.1126/science.1062872

Hahn, A., & Gawronski, B. (2014). Do implicit evaluations reflect unconscious attitudes? *Behavioral and Brain Sciences*, *37*, 28–29. doi:10.1017/S0140525X13000721

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834. doi:10.1037//0033-295X.108.4.814

Han, H., Glover, G. H., & Jeong, C., (2014). Cultural influences on the neural correlate of moral decision making processes. *Behavioral Brain Research, 259*, 215-228. doi:10.1016/j.bbr.2013.11.012

Hess, N. H., & Hagen, E. H. (2006). Psychological adaptations for assessing gossip veracity. *Human Nature*, *17*, 337-354. doi:10.1007/s12110-006-1013-z

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science, 345*, 1340-1343. doi:10.1126/science.1251560

Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology, 4,* 386.

doi:10.3389/fpsyg.2013.00386

Inbar, Y., Pizarro, D., & Cushman, F. (2012). Benefitting from misfortune: When harmless actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin, 38*, 52-62. doi:10.1177/0146167211430232

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology, 96*, 521-537. doi:10.1037/a0013779

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, 96*, 521-537.

doi:10.1037/0022-3514.89.6.899

Kagan, (1998). *Normative Ethics*. Westview Press: Boulder, CO.

Kant, I. (1785/1959). *Foundation of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis: Bobbs-Merrill.

Kawai, N., Kubo, K., & Kubo-Kawai, N. (2014). "Granny dumping": Acceptability of sacrificing the elderly in a simulated moral dilemma. *Japanese Psychological Research, 56*, 254-262.

Kenny, D. A., & DePaulo, B. M. (1993). Do people know how others view them? An empirical and theoretical account. *Psychological Bulletin, 114,* 145-161. doi:10.1037/0033-2909.114.1.145

Kohlberg, L. (1969). Stage and sequence: The cognitive–developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. (pp. 347–480). Chicago, IL: Rand McNally.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature 446*, 908-911. doi:10.1038/nature05631

Kreps, T. & Monin, B. (2014). Core values vs. common sense: Consequentialist views appear less rooted in morality. *Personality and Social Psychology Bulletin, 40*, 1529-1542. doi:10.1177/0146167214551154

Kundu, P., & Cummins, D. D. (2013). Morality and conformity: The Asch paradigm applied to moral decisions. *Social Influence*, *8*, 268-279. doi:10.1080./15534510.2012.727767

Laing, R. D., Phillipson, H., & Lee, A. R. (1966). *Interpersonal perception: A theory and a method of research*. Oxford: Springer.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93*, 234-249. doi:10.1037/0022-3514.93.2.234

Liu, B. S., & Ditto, P. H. (2013). What dilemma? Moral evaluation shapes factual belief. *Social Psychological and Personality Science, 4*, 316-323. doi:10.1177/1948550612456045

Lucas, B. L., & Galinsky, A. D. (2015). Is utilitarianism risky? How the same antecedents and mechanism produce both utilitarianism and risky choice. *Perspectives on Psychological Science*.

Lucas, B. L., & Livingstone, R. W. (2014). Feeling socially connected increases utilitarian choices in moral dilemmas. *Journal of Experimental Social Psychology, 53*, 1–4.

doi:10.1016/j.jesp.2014.01.011

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *TRENDS in Cognitive Sciences, 11*, 143-152. doi:10.1016/j.tics.2006.12.007

Mill, J. S. (1861/1998). *Utilitarianism.* In R. Crisp (Ed.), New York: Oxford University Press.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science, 19*, 549-57. doi:10.1111/j.1467-9280.2008.02122.x

Mumford, M. D., Zaccaro, S. J., Connelly, M. S., & Marks, M. A. (2000). Leadership skills: Conclusions and future directions. *The Leadership Quarterly*, *11*, 155-170.

Navarrete, C. D., McDonald, M. M., Mott, M. L. & Asher, B. (2011). Virtual morality: emotion and action in a simulated three-dimensional "trolley problem".
Emotion, 12, 364–370. doi:10.1037/a0025561

Nichols, S., Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition, 100,* 530-542. doi:10.1016/j.cognition.2005.07.005

Peeters, G. (1983). Relational and informational pattern in social cognition. In W. Doise & S. Moscovici (Eds.), *Current Issues in European Social Psychology* (201–237). Cambridge: Cambridge University Press.

Pinker, S. (2008, January 13). The moral instinct. *The New York Times*. Retrieved from http://www.nytimes.com

Pizarro, D. A. (2000). Nothing more than feelings? The role of emotions in moral judgment. *Journal for the Theory of Social Behavior, 30*, 355-375.

Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The Social Psychology of Morality: Exploring the Causes of Good and Evil* (pp. 91–108). Washington, DC: American Psychological Association.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717-731.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879-891. doi:10.3758/BRM.40.3.879

Prinz, J. (2006). The emotional basis of moral judgment. *Philosophical Explorations, 9*, 29-43.

Rockler, M. (2007). Presidential decision-making. *Philosophy Now*, *64*, 18-19.

Rosenberg, S., Nelson, C. & Vivekananthan, P. S., (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*, 283-294. doi:10.1037/h0026086

Royzerman, E.B., Landy, J. F., & Leeman, R. F. (2014). Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. *Cognitive Science, 39*, 1-28. 10.1111/cogs.12136

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of personality and social psychology*, *77*, 1004.

doi:10.1037/0022-3514.77.5.1004

Schnall, S., Roper, J., & Fessler, D. M. T. (2010). Elevation leads to altruism, above and beyond general positive affect. *Psychological Science, 21*, 315-320. doi:10.1177/0956797609359882

Schneid, E. D., Carlston, D. E., & Skowronski, J. J. (2015). Spontaneous evaluative inferences and their relationship to spontaneous trait inferences. *Journal of Personality and social Psychology, 108*, 681-696. doi:10.1037/a0039118

Singer, P. (2004). *The president of good and evil*. London: Granta.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file drawer. *Journal of Experimental Psychology: General, 143*, 534-547. doi:10.1037/a0033242

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, *88*, 895. doi:10.1037/0022-3514.88.6.895

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*, 164-182.

Sterba, J. P. (1988). *How to make people just*. Totowa, NJ: Rowman and Littlefield.

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

Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition, 119,* 454-458. doi:10.1016/j.cognition.2011.01.018

Szekely, R. D., Opre, A., & Miu, A. C. (2015). Religiosity enhances emotion and deontological choice in moral dilemmas. *Personality and Individual Differences, 79*, 104-109. doi:10.1016/j.paid.2015.01.036

Szekely, R. D. & Miu, A. C. (2014). Incidental emotions in moral dilemmas: The influence of emotion regulation. *Cognition & Emotion, 29*, 1-12. doi:10.1080/02699931.2014.895300

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology, 4*, 1-8. doi:10.3389/fpsyg.2013.00250

Tetlock, P. E. (2000). The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*, 853-870. doi:10.1037//0022-3514.78.5.853

Trémolière, B., & Bonnefon, J. F. (2014). Efficient kill–save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin, 40*, 923-930.

Trémolière, B., Kaminski, G., & Bonnefon, J. F. (2015). Intrasexual competition shapes men's anti-utilitarian moral decisions. *Evolutionary Psychological Science*, *1*, 18-22.

Uhlmann, E. L., Pizarro, D., A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making, 4*, 476-491.

Uhlmann, E. L., Zu, L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition, 126,* 326-334. doi:10.1016/j.cognition.2012.10.005

Vorauer, J. D., & Claude, S. (1998). Perceived versus actual transparency of goals in negotiation. *Personality and Social Psychology Bulletin, 24*, 371-385.

doi:10.1177/0146167298244004

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*, 548.

Whitlock, C. (2006, February 16). German court overturns law allowing hijacked airliners to be shot down. *Washington Post Foreign Service*. Retrieved from http://www.washingtonpost.com

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*, 395-412.

doi:10.1037/0022-3514.37.3.395

Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology, 67*, 222-232.

doi:10.1037/0022-3514.67.2.222

3.13 Interim Discussion

Contrary to Chapter 2 that focused on how people's idiosyncratic view influences person perception, Chapter 3 has found that different persons reach consensus regarding another person's personality independent of their own. Does that mean they are also aware how they are perceived by other people? On the one hand, people's insight regarding how others perceive them seems fairly limited. For example, people often fail to consider that others have less information than they do (Chambers et al., 2008; Epley et al., 2004). Therefore, people commonly base their meta-perceptions on self-perceptions (Chambers, Epley, Savitsky, & Windschitl, 2008; Kaplan, Santuzzi, & Ruscher, 2009; Kenny & DePaulo, 1993). On the other hand, there is ample evidence that people are aware how they are perceived by others (e.g., Carlson, Vazire, & Furr, 2011). For instance, meta-perception ratings particularly converge with social ratings when the underlying traits entail public behaviors (e.g., talkatively signals extraversion) rather than inner states (e.g., neurotic feelings, Vazire, 2010). Sharing one's dilemma judgment entails a clear public behavior, suggesting relative accuracy in meta-perceptions (Carlson et al., 2011, Carlson & Furr, 2009). Therefore, people may exhibit insight into how others perceive their warmth and competence following moral dilemma judgments. I argue that just as self-perception influences our impression of others, meta-perception may influence the way how we manage these impressions. If so, this opens up the possibility that people present themselves as warm or competent depending on which traits the situation favors.

Chapter 4 – The Strategic Moral Self: Self-Presentation Shapes Moral Dilemma Judgments

Abstract

Research has focused on the cognitive and affective processes underpinning dilemma judgments where causing harm maximizes outcomes. Yet, recent work indicates that lay perceivers infer the processes behind others' judgments, raising two new questions: whether decision-makers accurately anticipate the inferences perceivers draw from their judgments (i.e., meta-insight), and, whether decision-makers strategically modify judgments to present themselves favorably. Across seven studies, a) people correctly anticipated how their dilemma judgments would influence perceivers' ratings of their warmth and competence, though self-ratings differed (Studies 1-3), b) people strategically shifted public (but not private) dilemma judgments to present themselves as warm or competent depending on which traits the situation favored (Studies 4-6), and, c) self-presentation strategies augmented perceptions of the weaker trait implied by their judgment (Study 7). These res8ults suggest that moral dilemma judgments arise out of more than just basic cognitive and affective processes; complex social considerations causally contribute to dilemma decision-making.

Keywords: moral dilemmas, social judgment, social perception, self-perception, meta-perception

The Strategic Moral Self: Self-Presentation Shapes Moral Dilemma Judgments

During the Second World War, Alan Turing and his team cracked the Enigma Code encrypting German war communications. Soon, British High Command discovered an impending attack on Coventry—but taking countermeasures would reveal the decryption (Winterbotham, 1974). Thus, they faced a moral dilemma: allow the deadly raid to proceed and continue intercepting German communications, or deploy lifesaving countermeasures and blind themselves to future attack. Ultimately, the Allies allowed the attack to proceed. Lives were lost, but some analysts suggest this decision expedited the war's conclusion (Copeland, 2012). The moral judgment literature suggests that such decisions reflect a tension between basic affective processes rejecting harm and cognitive evaluations of outcomes allowing harm (Greene, 2014). But is it possible that self-presentation also factored in? The British High Command may have considered how their allies would react upon learning they threw away a tool for victory to prevent one deadly, but relatively modest, raid.

Moral dilemmas typically entail considering whether to accept harm to prevent even greater catastrophe. Philosophers originally developed such dilemmas to illustrate a distinction between killing someone as the means of saving others versus as a side effect of doing so (Foot, 1967), but subsequent theorists have largely described them as illustrating a conflict between deontological and utilitarian philosophy (e.g., Greene et al., 2001). The dual process model suggests that affective reactions to harm underlie decisions to reject harm, whereas cognitive evaluations of outcomes underlie decisions to accept harm to maximize outcomes (Greene, 2014). Other theorists have described these as processes in terms of basic cognitive architecture for decision-making (Cushman, 2013; Crockett, 2013), or heuristic adherence to moral rules (Sunstein, 2005). Notably, all such existing models focus on relatively basic, non-social processing.

Yet, Haidt (2001) argued that moral judgments are intrinsically social, and communicate important information about the speaker. Indeed, recent work indicates that lay perceivers view decision-makers who reject harm (upholding deontology) as warmer, more moral, more trustworthy, more empathic, and more emotional than decision-makers who accept harm (upholding utilitarianism), whom perceivers view as more competent and logical, with consequences for hiring decisions (Everett, Pizarro, and Crocket, 2016; Kreps & Monin, 2015; Rom, Weiss, & Conway, 2017; Uhlmann, Zhu, and Tannenbaum, 2013).[17] Moreover, social pressure can influence dilemma judgments (Bostyn & Roets, 2016, Kundu & Cummins, 2012; Lucas & Livingstone, 2014). Such findings raise the question of whether people have meta-insight into how their dilemma judgments make them appear in the eyes of others, and whether decision-makers *strategically* adjust dilemma judgments to create desired social impressions. If so, this would provide the first evidence to our knowledge that higher-order processes causally influence judgments, suggesting dilemma decisions do not merely reflect the operation of basic affective and cognitive processes.

4.1 Moral Dilemma Judgments: Basic vs. Social Processes

Moral dilemmas originated as philosophical thought experiments, including the famous trolley dilemma where decision-makers could redirect a runaway trolley so it kills one person instead of five (Foot, 1967). According to Greene and colleagues (2001), refusing to cause harm to save others qualifies as a ‘characteristically deontological’ decision, because in deontological ethics the morality of action primarily hinges on its intrinsic nature (Kant, 1785/1959). Conversely, causing harm by redirecting the trolley saves five people, thereby qualifying as a

[17] Deontological dilemma judgments appear to convey both warmth and morality (Rom et al., 2017). Although these constructs can be disentangled (e.g., Brambilla et al., 2011), in the present context they happen to covary substantially. It may be that different aspects of deontological decisions influence these perceptions (e.g., whether they accord with moral rules; whether they suggest emotional processing), but these aspects overlap in the current paradigm. We focus primarily on perceptions of warmth, which roughly corresponds to the affective processing postulated by the dual process model, and relegated findings regarding morality the supplement. Future work should disentangle warmth trait perceptions from moral character evaluations.

'characteristically utilitarian' decision, because in utilitarian ethics the morality of an action primarily hinges on its outcomes (Mill, 1861/1998).[18] Note that utilitarian philosophy technically entails impartial maximization of the greater good, which represents a subset of the broader concept of consequentialism, which advocates for outcome-focused decision-making more generally. We do not wish to imply that making a judgment consistent with utilitarianism renders one a utilitarian—it need not (e.g., Kahane, 2015)—but rather we use the term 'utilitarian' in the simpler senses that such judgments a) objectively maximize overall outcomes, b) appear to often entail ordinary cost-benefit reasoning, and c) utilitarian/consequentialist philosophers generally approve of such judgments (see Amit & Greene, 2010).

Although dilemmas originated in philosophy, research in psychology, neuroscience, and experimental philosophy has aimed to clarify the psychological mechanisms driving dilemma judgments. Most prominent among these is the dual process model, which postulates that basic affective and cognitive processes drive dilemma judgments (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Other theorists have argued judgments reflect decision-making systems focused on immediate action versus long-range goals (Cushman, 2013; Crockett, 2013), heuristic adherence to moral rules (Sunstein, 2005), or the application of innate moral grammar (Mikhail, 2007). We do not aim to adjudicate between these various claims, nor do we dispute the contribution of such processes. Rather, we simply note that these models focus on basic, nonsocial processes.

[18] Following Greene and colleagues (2001), we use the term 'characteristically' deontological/utilitarian, because there are many variants of each theory that do not all agree. Nonetheless, this terminology is widely employed currently, and so we follow in this terminological tradition despite its limitations. Note that we are not arguing that making a given dilemma decision implies that decision-makers ascribe to abstract philosophical commitments. Rather, we argue simply that 'utilitarian' judgments qualify as such because they tend to maximize outcomes, regardless of decision-makers' philosophical commitments. Just as one need not be Italian to cook an Italian meal, accepting outcome-maximizing harm on a dilemma does not make one a utilitarian. Hence, these terms reflect only to the content of judgments, rather than the qualities of judges (see Amit & Greene, 2012).

Research has largely ignored the possibility that higher-order sophisticated social processes might causally contribute to dilemma judgments. Yet, morality appears intrinsically social (Haidt, 2001), and most real-world moral judgments involve publicly communicating with others (e.g., Hofmann, Wisneski, Brandt, & Skitka, 2014). We expect the same is true of dilemma judgments. Although the best-known dilemmas are hypothetical (such as the trolley dilemma), many real-world decisions entail causing harm to improve overall outcomes (e.g., launching airstrikes in Syria to prevent ISIS from gaining momentum, punishing naughty children to improve future behavior, imposing fines to prevent speeding). As decisions in such cases align with either deontological or utilitarian ethical positions, they correspond to real world moral dilemmas. Moreover, lay decision-makers employ verbal arguments that align with deontological and utilitarian ethical positions (Kreps & Monin, 2014). Hence, social consideration of dilemma judgments is not restricted to responses to hypothetical scenarios, but forms an ordinary part of communication about common moral situations.

Kreps and Monin (2014) examined deontological and utilitarian arguments in speeches by Presidents Clinton and Bush, among other politicians. Lay perceivers viewed speakers as moralizing more when they framed arguments in terms of deontology rather than utilitarianism. These findings align with work on hypothetical dilemma decisions: perceivers rated and treated decision-makers who rejected harm (upholding deontology) as more trustworthy than decision-makers who accept harm (upholding utilitarianism, Everett et al., 2016), as well as more moral, more empathic, and less pragmatic than harm-accepting decision-makers (Uhlmann et al., 2013). Likewise, Rom and colleagues (2016) found that lay people appear to intuit the dual process model: they rated targets who rejected harm as relatively warm, and inferred that such judgments were driven by emotion. Conversely, perceivers rated targets who accepted harm as relatively competent, and inferred that such judgments were driven by cognitive deliberation.[19]

[19] If the dual-process model is correct, responses to classic moral dilemmas do not reflect the degree to which decision-makers experience affective reactions or engage in cognition in an absolute sense. If

Moreover, perceivers preferred harm-rejecting decision-makers for social roles prioritizing warmth, such as social partners or their child's doctor, but preferred harm-accepting decision-makers for roles prioritizing competence, such as hospital administration (Everett et al., 2016, Rom et al., 2017). Hence, decision-makers face a warmth/competence tradeoff when presenting their decision to others. The current work examines whether decision-makers are aware of this trade-off, and whether they strategically adjust their decisions to present themselves favorably.

4.2 Meta-Perceptions Regarding Dilemma Judgments

We propose that lay perceivers hold fairly accurate meta-perceptions into how others will view them based on their dilemma decision. People care deeply about their moral reputation (Aquino & Reed, 2002; Krebs, 2011; Everett et al., 2016) and the moral reputations of others (Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Goodwin, Piazza, & Rozin, 2014). Clearly, the research described above on perceptions of decision-makers indicate that dilemma decisions can affect moral reputation, suggesting that people should be attuned to what messages their judgments convey. Moreover, past work suggests that people can be reasonably accurate when gauging how others perceive them. For example, narcissists appear aware that others view them less positively than they view themselves (Carlson, Vazire, & Furr, 2011, Carlson & Furr, 2009). Self- and social-ratings particularly converge when the underlying traits entail public behaviors (e.g., loquaciousness signals extraversion) rather than inner states (e.g., neurotic feelings,

classic moral dilemmas place affect and cognition in conflict, and ultimately judges may only choose one option, then judgments reflect the *relative* strength of each process. For example, accepting harm that maximizes outcomes may occur either due to strong cognition coupled with strong but slightly weaker affect, or weak cognition coupled with weaker affect. Hence, a judgment to accept causing harm does not reveal whether the judge experienced strong or weak affect—only that cognition outweighed whatever degree of affect they experienced. Nor does such a judgment guarantee that the judge engaged in strong cognition—only that whatever cognition they engaged in outweighed their affective experience. Some people may engage in extensive affect and cognition, whereas others engage in little of either. In order to estimate each processes independently, it is necessary to use a technique such as process dissociation (see Conway & Gawronski, 2013). However, in the current work we are not interested in the actual processes underlying dilemma judgments so much as lay perceptions of these processes. To that end, lay people, like many researchers, equate harm avoidance judgments with strong affect and harm acceptance judgments with strong cognition. This intuition is effective as a rough heuristic, so long as researchers recognize that it does not accurately describe moral dilemma processing.

Vazire, 2010). Sharing one's dilemma judgment entails a clear public behavior, suggesting relative accuracy in meta-perceptions.

However, other research casts doubt on the possibility of accurate dilemma meta-perceptions in dilemma research. Besides public expression, dilemma judgments entail intrapsychic aspects such as emotional reactions, perceptions of conflict, and so on (e.g., Kruger & Gilovich, 2004; Anderson & Ross, 1984; Pronin, 2008; Winkielman & Schwarz, 2001). Decision-makers hold privileged knowledge of their experience of these inner states. People often fail to consider that others have access to less information than they do (Chambers, Epley, Savitsky, & Windschitl, 2008). Whereas egocentric perspectives come to mind easily, adjusting away from egocentricity is difficult (Epley, Keysar, Van Boven, & Gilovich, 2004). Thus, meta-perceptions are often biased by self-understanding (Chambers et al., 2008; Kaplan, Santuzzi, & Ruscher, 2009; Kenny & DePaulo, 1993). Moreover, people are motivated to view themselves positively in the moral domain (Epley & Dunning, 2000) much like non-moral domains (e.g., Dunning & McElwee, 1995), and can rationalize either dilemma decision in self-flattering ways (Uhlmann, Pizarro, Tannenbaum, & Ditto, 2009; Liu & Ditto, 2014). Thus, people may well judge themselves as high in both warmth and competence regardless of their dilemma decision—and may expect others to agree with this flattering self-assessment.

If decision-makers erroneously base meta-perceptions on self-perceptions, meta-perception ratings should converge with self-ratings and diverge from ratings of others following the same judgment—that is, people may believe they come across as both warm and competent regardless of their dilemma decision, whereas they view others' decisions as reflecting a warmth/competence trade-off. Conversely, if people have accurate meta-insight into how others perceive them, meta-perception ratings should converge with other ratings and diverge from self-ratings—that is, people may privately believe they are warm and competent regardless of dilemma decision, yet expect others to rate them according to the same warmth/competence tradeoff implied by others' judgments. We contrasted these predictions empirically.

4.3 Strategic Self-Presentation in Dilemma Judgments

If people evince accurate meta-insight into what their dilemma decision conveys, this raises the possibility that they strategically adjust such decisions to present themselves favorably. There are potential upsides and downsides to selecting each dilemma judgment, as the precise cause of others' dilemma decisions appear ambiguous. Upholding utilitarianism by accepting outcome-maximizing harm amounts to bloodying one's hands for the sake of the community. Such bold and brutal action may convey either competent leadership (Lucas & Galinksy, 2015) or a callous disregard for causing harm—as in psychopathy (Bartels & Pizarro, 2011) or low empathy (Gleichgerrcht & Young, 2013). Conversely, rejecting harm (upholding deontology) may convey either a warm concern for others and/or principled respect for life and/or trustworthiness (Everett et al., 2016; Kreps & Monin, 2015; Rom et al., 2017), or suggest incompetent paralysis when the situation demands bold action (Gold et al 2015; Gawronksi et al., 2015). Hence, in some circumstances it may be preferable to risk appearing incompetent in order to convey warmth, trustworthiness, and respect for life; in other situations, it may be preferable to risk appearing cold and callous in order to convey decisive competence and leadership.

People care deeply about presenting themselves favorably. They tailor their public images in various domains to the perceived values and preferences of important others (Reis & Gruzen, 1976; von Baeyer, Sherk, & Zanna, 1981; Leary & Kowalski, 1990; Leary, 1995). People change social roles over time, and social roles carry expectations regarding how individuals who occupy those roles ought to behave (Sarbin & Allen, 1968). Hence, people often flexibly present themselves to conform to different social role expectations (Leary, Robertson, Barnes, & Miller, 1986; Leary, 1989). Indeed, Everett and colleagues (2016) argued that deontological dilemma judgments may operate as a reputation-management mechanism to present oneself as a trustworthy social interaction partner by demonstrating respect for others autonomy and wishes (see also Bostyn & Roets, 2016).

Accordingly, previous work demonstrates that social situations influence dilemma responses. In a modification of the Asch conformity paradigm, Kundu and Cummins (2012) asked participants whether they would accept or reject outcome-maximizing harm after a series of confederates gave a particular answer. They found evidence for conformity pressure: participants were more likely to give answers consistent with those of the confederates. Bostyn and Roets (2016) conducted a similar study, and argued that conformity pressure was stronger for harm rejection (upholding deontology) than harm acceptance (upholding utilitarianism). However, Lucas and Livingstone (2014) found that participants who socially connected with others before completing dilemmas after were more willing to accept harm (upholding utilitarianism). It may be that resolving dilemmas in front of strangers motivated participants to skew towards deontological answers so as to avoid appearing immoral—after all, research suggests that moral traits appear especially important when forming first impressions (Brambilla et al., 2011; Goodwin et al., 2014), and that warmth may also be important when forming first impressions (Fiske, Cuddy, & Glick, 2007). Conversely, when participants have an opportunity to establish warmth or morality through social interactions, they may have felt free to demonstrate other qualities, such as competence. These findings suggest that context may shift whether accepting or rejecting harm seems to be the optimal answer. If participants strategically adjust dilemma judgments, their perception of expectations should vary depending on whether the circumstances appear to prioritize warmth over competence, and their public (but not private) dilemma answers should track such expectations.

4.4 Overview

Across seven studies, we investigated whether people hold accurate meta-perceptions regarding how others view them based on their dilemma judgments, and whether they strategically modify such judgments to present themselves favorably. First, we examined whether people have accurate meta-insight into the warmth and competence ratings others infer from their dilemma judgments by comparing warmth and competence ratings of others, the self, and meta-perceptions of the self (Studies 1-3). Second, we tested whether people shift public (but not private) dilemma judgments depending on whether warmth or competence is favored in a given situation (Studies 4-6). Third, we investigated whether people can use communication strategies to offset the weaker trait implied by their judgment—whether people who accept harm can come across as warm, and people who reject harm can come across as competent (Study 7). Across all studies, we disclose all measures, manipulations, and exclusions, as well as the method of determining the final sample size. In none of the studies data collection was continued after data analysis.

4.5 Study 1

Study 1 examined the accuracy of participants' meta-perceptions (i.e., meta-accuracy, Anderson, Ames, & Gosling, 2008) following moral dilemma judgments. We randomly assigned participants to one of three conditions: participants either made a dilemma judgment themselves (self and meta-perception condition) or read about another persons' dilemma judgment (other condition). Then, participants in the self-condition rated their own warmth and competence, those in the other condition rated the others' warmth and competence, and those in the meta-perception condition rated how they believed others would view their warmth and competence. Hence, the design was a 3 (target: self vs. other vs. meta-perceptions) × 2 (decision: harm rejection vs. acceptance) × 2 (personality dimension: warmth vs. competence) quasi-experimental design (as participants were free to make either dilemma judgment themselves) with the first two factors between-subjects and the third within-subjects.

Given that people tend to view themselves positively in the moral domain (Epley &

Dunning, 2000), and have access to internal perceptions of conflict between response options, we expected participants in the self condition would rate themselves high on both warmth and competence, regardless of their dilemma decision. We expected participants in the other condition to replicate the patterns demonstrated by Rom and colleagues (2016): they should rate targets who rejected causing harm as warmer but less competent than targets who accepted causing outcome-maximizing harm. Most importantly, we predicted that participants' meta-perception condition would exhibit meta-accuracy, by anticipating that others would rate them using the same warmth/competence tradeoff (depending on dilemma decision) as participants in the other condition, rather than the uniformly high warmth and competence ratings participants privately make about themselves.

4.5.1 Method

Participants. We recruited 200 American participants (134 males, 66 females, M_{age} = 30.63, *SD* = 8.92) via Mechanical Turk, who received $0.25, aiming for ~50 per between-subjects condition, although actual responses varied substantially ($n_{\text{self_harm_rejection}}$ = 14; $n_{\text{self\_harm\_acceptance}}$ = 30; $n_{\text{other\_harm\_rejection}}$ = 54; $n_{\text{other\_harm\_acceptance}}$ = 46; $n_{\text{meta\_harm\_rejection}}$ = 14; $n_{\text{meta\_harm\_acceptance}}$ = 42). First, we randomly assigned participants to either learn about Brad's ostensible judgment or to make a judgment themselves. Next, we randomly assigned half of participants in the self-dilemma-judgment condition to rate themselves on warmth and competence, and half to rate themselves as they expected others would (meta-perceptions).[20] A Levene's test of equality of error variances revealed that in homogeneity was not violated for warmth, *F*(5,194) = 1.88, *p* = .100, but for competence, *F*(5,194) = 3.15, *p* = .009.We excluded no one. Although we did not conduct a priori power analyses, we felt confident that this design provided reasonable power based on past work (Rom et al., 2017). Indeed, a post hoc power

[20] We acknowledge that this two-stage random assignment is suboptimal because it led to uneven cell sizes, which is one reason we increased the sample size in Study 2.

analysis using GPower (Faul, Erdfelder, Lang, & Buchner, 2007) for a fixed-effects between-within design where η_p^2 = .10, *N* = 200, α = .05, and the correlation between repeated measures was *r* = .33, suggested that we had ~99% power to detect the obtained interaction.

Procedure. All participants read the widely-employed crying baby dilemma (e.g., Conway & Gawronski, 2013), where the actor must decide whether to smother a baby to prevent its cries from alerting murderous soldiers hunting for other townspeople in hiding. Participants in the self and meta-perception conditions then selected either *yes, this action is appropriate* or *no, this action is not appropriate* (following Greene et al., 2001). Participants in the other condition viewed a photo of a university student named Brad, then learned that Brad had selected either one or the other of these responses (following Rom et al., 2017). Then, participants completed measures of warmth and competence using items adapted from Fiske, Cuddy, Glick and Xu (2002).

Depending on condition, participants either rated themselves, Brad, or indicated how they thought others would rate them following their decision (meta-perception). Specifically, those in the meta-perception condition read:

> *Now take a moment to imagine that another person saw the judgment you made.*
>
> *Based on that information, what would they think about you? From their perspective how well do you think they would say each trait describes you? THEY would think you are…*

Participants indicated how well four warmth traits (*warm, good-natured, tolerant, sincere*) and five competence traits (*competent, confident, independent, competitive, intelligent*) described the target on 7-point scales anchored at 1 (*not at all*) and 7 (*very much*). We averaged judgments into composites of warmth (α = .91) and competence (α = .87), which were modestly correlated (*r* = .33). Item order was randomized for each participant. For exploratory reasons, we also included the single item *moral*, consistent with Rom and colleagues (2016).

Some researchers have argued that morality and warmth are distinguishable constructs (Brambilla, et al. 2011; Goodwin et al., 2014). We find these arguments persuasive—used car salesmen that evince warm sociability should not be trusted, whereas a cold and dispassionate judge who sentences criminals may nonetheless appear moral. Nonetheless, it may be that these constructs align more in some contexts than others. Hence, we empirically examined how well these constructs dissociated in the current studies using five strategies.

First, we noted that the item *moral* consistently correlated highly with the warmth composite measure, ~*r* = .75, consistent with Rom and colleagues (2016). Second, we noted that the item *moral* varied across conditions in the same manner as the warmth composite on all studies (see Supplementary analysis). Now, it remains possible that these findings simply reflect the fact that some items in the warmth composite—such as *sincerity*—assess perceptions morality instead of warmth. Therefore, third, we conducted factor analyses (principle axis factoring with oblimin rotation) for all studies assessing warmth and morality (see Table S1 in supplementary material). In each case, all warmth items loaded together with the item *moral* onto a single factor, whereas all competence items loaded onto a separate factor. A couple of items occasionally loaded well on both factors—*confident*, *tolerant*, *competent*, and *intelligent*—but these dual loadings each occurred only once, and did not replicate across the other studies. Fourth, we conducted follow-up analyses for each study using only the single items warmth and competent instead of the composite measures; findings were very similar (find an example for Study 1 in the supplementary material). Fifth, we conducted follow-up analyses for each study using an alternative warmth score based on two items (*warm, good-natured*), and an alternative morality score based on three items (*sincere, tolerant, morality*),[21] as well as a

[21] We thank an anonymous reviewer for this suggestion.

combined warmth/morality score including all warmth items plus the item morality. In each case, the pattern of findings remained very similar to the patterns presented below.

These findings suggest that in the context of dilemma perceptions, participants may find it difficult to disentangle warmth and morality. After all, perceivers may find it ambiguous whether a given deontological judgment reflects affective processing or adherence to moral rules. Alternatively, it may be that the particular items presented in this scale underestimate the difference between these constructs. Either way, the current paradigm was not designed to distinguish between warmth and morality. Indeed, these analyses suggest it may even be warranted to include the item moral in the warmth composite measure. Nonetheless, in recognition of the important theoretical distinction between warmth and morality (Brambilla & Leach, 2011; Goodwin et al., 2014) and to remain consistent with Rom and colleagues (2016), we decided to treat the item morality as a separate construct. Given that the current focus was on contrasting perceptions of warmth and competence, and the similarity between the patterns of warmth and morality, we decided to relegate the morality findings to the supplementary material.

4.5.2 Results

We submitted ratings to a 3 (target: self vs. other vs. meta-perceptions) × 2 (decision: harm rejection vs. acceptance) × 2 (personality dimension: warmth vs. competence) repeated measures ANOVA with the first two factors between and the last factor within subjects (see Figure 1). We conducted Levene's tests to examine homogeneity of variance assumptions. This assumption was not violated for warmth, F(5,194) = 1.88, p = .100, but was violated for competence, F(5,194) = 3.15, p = .009. Therefore, to supplement the main analysis in the text, we also conducted non-parametric Kruskal-Wallis and Mann-Whittney tests (see Supplement), which are more robust to violations of homogeneity of variance (Tomarken & Serlin, 1986; Kruskal & Wallis, 1954; Mann & Whittney, 1947). The results of these tests largely corroborated the conclusions of the main analyses presented here. There was a main effect of target: participants gave higher ratings overall in the self (*M* = 5.18, *SD* = 1.12) than other (*M* = 4.64, *SD* = .90), or meta-perception conditions (*M* = 4.28, *SD* = 1.04), *F*(2, 194) = 8.47, $p < .001$, η_p^2 = .08. There was also a main effect of decision: participants rated targets who rejected harm, upholding deontology, higher overall (*M* = 4.86, *SD* = 1.10), than targets who accepted harm, upholding utilitarianism (*M* = 4.51, *SD* = 1.03), *F*(2, 194) = 8.32, *p* = .004, η_p^2 = .04. There was no main effect of personality dimension, *F*(2, 194) = 1.75, *p* =.18, η_p^2 = .01. These main effects were qualified by a significant two-way interaction between target decision and personality measure, *F*(1, 194) = 45.65, $p < .001$, η_p^2 = .19, and a marginal interaction between target and personality measure, *F*(2, 194) = 3.03, *p* = .050, η_p^2 = .03, 95%, whereas the interaction between target and decision was not significant, *F*(2, 194) = 1.55, *p* = .214, η_p^2 = .02. Moreover, the three-way interaction was significant, *F*(2, 194) = 11.14, $p < .001$, η_p^2 = .10.

We decomposed these interactions by examining post-hoc tests within each condition. As predicted, participants in the self-condition rated themselves equally high on warmth when they rejected (*M* = 5.69, *SD* = 1.24) or accepted (*M* = 5.12, *SD* = 1.40) causing harm, *F*(1,194) = 2.70, p =.102, η_p^2 = .01, and equally competent when they rejected (*M* = 5.50, *SD* = 1.17) versus accepted causing harm (*M* = 4.85, *SD* = 1.94), *F*(1,194) = 3.60, *p* = .059, η_p^2 = .03. However, participants in the other-condition replicated the predicted warmth/competence tradeoff found previously: Participants rated Brad higher on warmth when he rejected (*M* = 5.00, *SD* = 1.19), than when he accepted causing outcome-maximizing harm (*M* = 4.03, *SD* = .99), *F*(1, 194) = 15.57, $p < .001$, η_p^2 =.07. Conversely, they rated Brad as higher in competence when he accepted (*M* = 5.16, *SD* = 1.16), rather than rejected causing outcome-maximizing harm (*M* = 3.36, *SD* = 1.31),-*F*(1, 194) = 11.67, $p < .001$, η_p^2 =.06.

Crucially, participants in the meta-perception-condition evinced the same warmth/competence tradeoff as participants in the other-condition: When participants rejected harm they inferred others would perceive them as warmer (M = 5.16, SD = 1.59) than when they accepted causing outcome-maximizing harm (M = 3.36, SD = 1.31), F(1, 194) = 22.95, p < .001, η_p^2 =.10. In contrast, when they accepted such harm, they inferred that others would perceive them as equally competent (M = 4.89, SD = 1.10) than when they rejected such harm (M = 4.38, SD = 1.46), F(1, 194) = 2.32, p = .129, η_p^2 =.01.

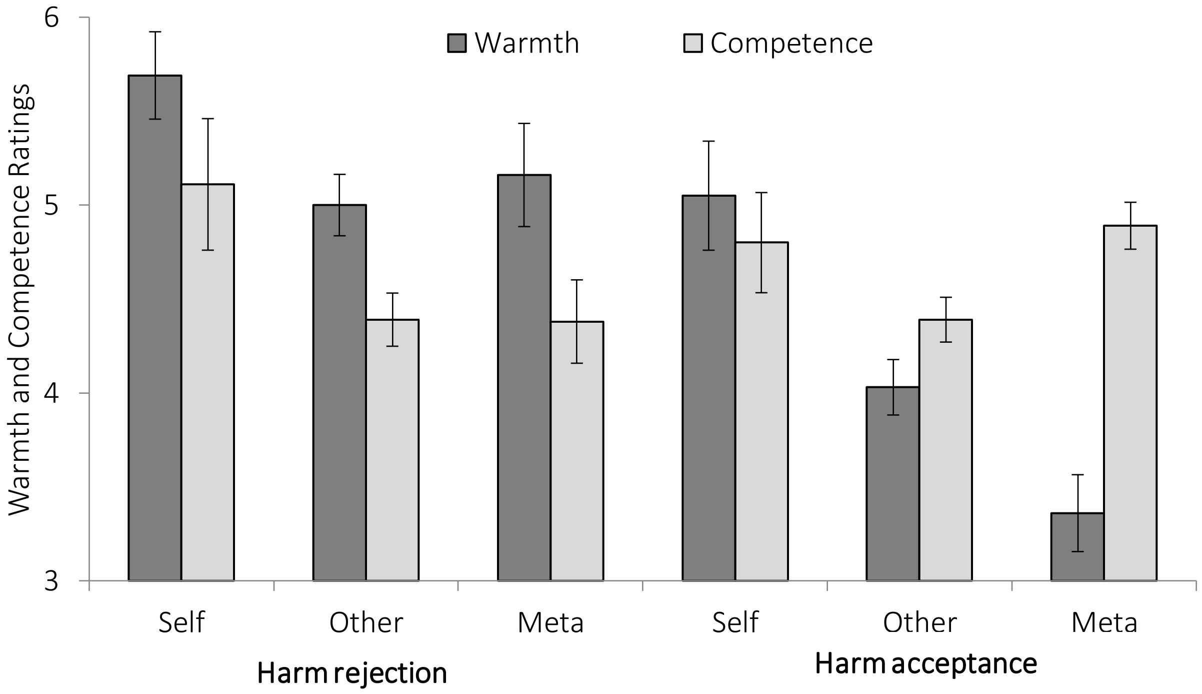


Figure 1. Participants' self, target, and meta-perception warmth and competence ratings when they or the target rejected causing harm to maximize outcomes (upholding deontology), or accepted such harm (upholding utilitarianism), Study 1. Error bars reflect standard errors.

4.5.3 Discussion

These findings suggest that participants have accurate meta-insight regarding the inferences others will draw about their personality from their dilemma judgments. Privately, participants rated themselves equally high on warmth and competence regardless of their dilemma decision. However, in the meta-perception condition they expected others to rate them similar to how they rated others: just as participants viewed targets who rejected causing harm as warmer and less competent than targets who accepted causing harm, they expected that others would rate them as warmer and less competent when they rejected vs. accepted causing harm themselves. To our knowledge, this is the first evidence that participants are aware of the impression their dilemma judgments convey to others.

However, our quasi-experimental design suffered from the limitation of nonrandom assignment: participants in the self and meta-perception conditions freely choose which dilemma decision to make. Hence, it remains possible that our meta-perception results reflect the general psychology of people who made a specific decision, rather than inferences regarding that decision per se. Even though this interpretation seems unlikely give the null effect in the private self-rating condition, we aimed to resolve this confound in Study 2.

4.6 Study 2

Study 2 replicated the meta-perception condition from Study 1, together with a *communication error* condition where participants imagined that others erroneously learned they made the dilemma judgment opposite to the one they truly made. This design allowed us to test whether meta-perceptions in Study 1 would hold for decisions that participants personally disagreed with. We expected that warmth and competence meta-perceptions would track the decision others believed participants made (harm rejection: higher warmth than competence, harm acceptance: higher competence than warmth), rather than the decision participants actually made.

4.6.1 Method

Participants. To increase confidence in the effects and address the uneven cell sizes in Study 1, we decided to approximately double the sample size and employ more traditional randomization procedures. We recruited 397 American participants via Mechanical Turk, who received $0.25. We excluded 24 participants who completed less than 50% of dependent measures, leaving a final sample of 373 (244 males, 123 females, 6 unreported, M_{age} = 30.49, *SD* = 9.89. Participants were randomly assigned to the correct versus error condition ($n_{\text{correct\_harm\_rejection}}$= 32; $n_{\text{correct\_harm\_acceptance}}$ = 157; $n_{\text{error\_harm\_rejection}}$ = 45; $n_{\text{error\_harm\_acceptance}}$ = 139). In both conditions, many more people accepted than rejected harm, but due to the communication error these ratios appear to flip. A Levene's test of equality of error variances revealed that homogeneity was neither violated for warmth, *F*(3,369) = 2.50, *p* = .059, nor for competence, *F*(3,369) = 1.31, *p* = .271. GPower suggested we had ~99% post-hoc power to detect the obtained interaction with this sample size.

Procedure. Each participant read the crying baby dilemma from Study 1, and selected one of the two dilemma responses. Then we randomly assigned them to the *correct communication* or *communication error* condition. Participants in the correct communication condition imagined that others correctly learned which dilemma decision they made, as in Study 1. Participants in the communication error condition imagined that others erroneously learned they made the dilemma decision opposite to their real decision. Specifically, participants read:

> *Now take a moment to imagine that another person learned about the judgment you made. As often happens, misinformation got out and this other person thinks you chose: Yes, harm is appropriate [No harm is not appropriate]. Based on the information that you would [not] SMOTHER the baby, what would this person think of you? From their perspective how well do you think they would say each trait describes you? THEY would think you are…*

Participants indicated how they believed others would perceive them on the same warmth (α = .89), competence (α =.87), and morality items as Study 1. This resulted in a 2 (communication: correct vs. error) × 2 (decision: harm rejection vs. acceptance) × 2 (dimension:

warmth vs. competence) quasi-experimental design, as participants could not be randomly assigned to make a particular judgment. Consistent with Study 1 and past work (Rom et al., 2017), warmth and competence correlated moderately (r = .40), whereas morality correlated highly with warmth (r = .87) and less with competence (r = .38). Morality yielded results similar to warmth across condition (replicating previous work, Rom et al., 2017) but was not focus of the current manuscript, so we again relegated it to the supplement.

4.6.2 Results

We submitted warmth and competence ratings to a 2 (communication: correct vs. error) × 2 (decision: harm rejection vs. acceptance) × 2 (dimension: warmth vs. competence) repeated measures ANOVA (see Figure 2) with the first two factors between-subjects and the last factor within-subjects. There was a main effect of communication: participants gave higher personality ratings overall in the correct communication (M = 4.25, SD = 1.26) than communication error condition (M = 3.98, SD = 1.35), F(1, 369) = 17.51, p < .001, η_p^2 = .05. There was no main effect of decision, F(1, 369) = 0.46, p = .499, η_p^2 = .001, but there was a main effect of personality dimension: participants gave lower warmth (M = 3.89, SD = 1.81) than competence ratings overall (M = 4.35, SD = 1.47), F(1, 369) = 7.30, p = .007, η_p^2 = .02. In addition, there were significant 2-way interactions between communication and personality dimension, F(1, 369) = 19.26, p < .001, η_p^2 = .05, and between decision and personality dimension, F(1, 369) = 4.43, p = .036, η_p^2 = .01, though not between communication and personality dimension, F(1, 369) = 2.25, p = .134, η_p^2 = .01. More importantly, we obtained the expected three-way interaction, F(1, 369) = 49.02, p < .001, η_p^2 = .12.

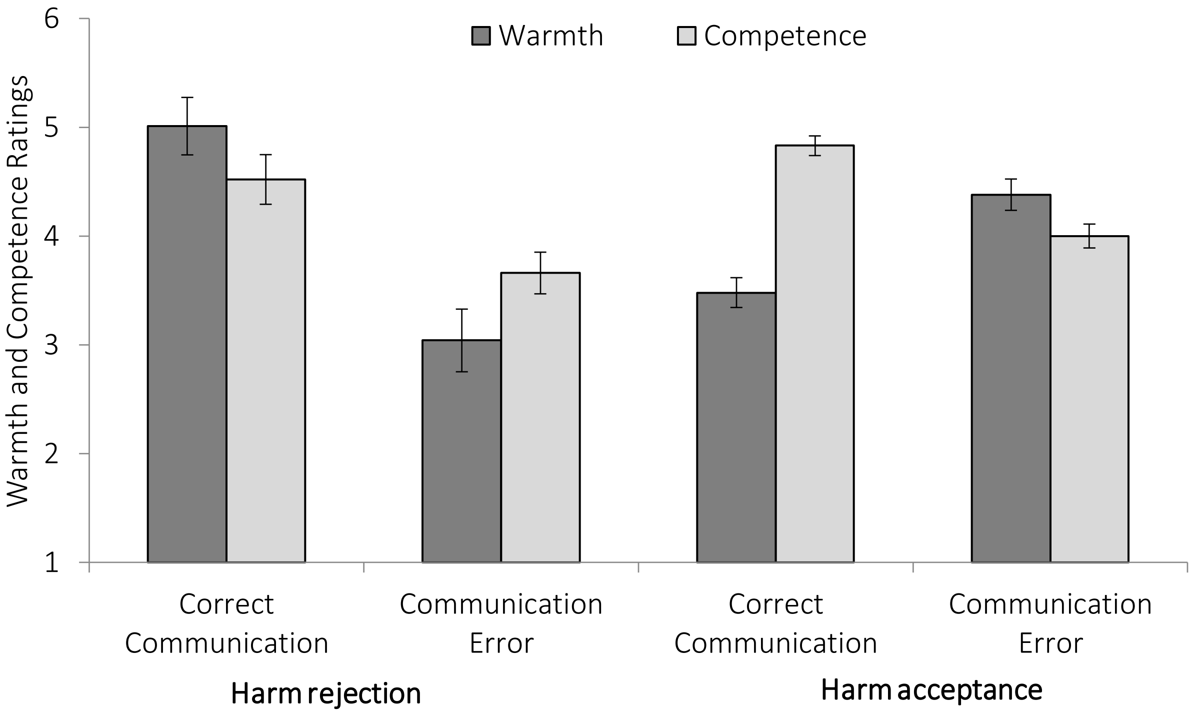


Figure 2. Warmth and competence meta-perceptions when participants rejected outcome-maximizing harm (upholding deontology), or accepted such harm (upholding utilitarianism), and imagined others correctly learned their judgment (correct communication condition) or erroneously believed they made the opposite judgement (communication error condition), Study 2. Error bars reflect standard errors.

Post-hoc contrasts largely replicated Study 1 in the correct communication condition: Participants expected that others would rate them as warmer when they rejected harm, upholding deontology (M = 5.06, SD = 1.49) than accepted causing harm, upholding utilitarianism (M = 3.40, SD = 1.68), F(1, 182) = 19.70, p < .001, η_p^2 = .10. Results for competence trended in the expected direction, but did not reach significance: participants expected that others would rate them as similarly competent when they rejected (M = 4.51, SD = 1.19), rather than accepted, causing harm (M = 4.82, SD = 1.19), F(1, 187) = 1.91, p = .168, η_p^2 = .01. Participants in the error communications condition showed the opposite pattern.

Participants expected that others would rate them as less warm when they rejected (*M* = 2.94, *SD* = 1.93) rather than accepted causing harm (*M* = 4.42, *SD* = 1.68), *F*(1, 369) = 22.28, *p* < .001, η_p^2 = .11. Again, ratings for competence trended nonsignificantly in the expected direction: participants expected others to rate them similarly on competence when they rejected (*M* = 3.66, *SD* = 1.40), versus accepted causing harm (*M* = 4.00, *SD* = 1.29), *F*(1, 369) = 2.25, *p* = .135, η_p^2 = .01.

4.6.3 Discussion

Study 2 replicated the findings from Study 1 in the correct communication condition: participants who rejected harm (upholding deontology) inferred that others would perceive them as relatively warmer but (nonsignificantly) less competent, whereas participants who accepted harm (upholding utilitarianism) inferred that others may perceive them as (nonsignificantly) more competent but less warm. Moreover, these meta-perception ratings flipped when participants imagined that a communication error occurred, and others erroneously believed they made the judgment opposite to the judgment they actually made. Hence, meta-perceptions tracked the information available to others, rather than reflecting the judgments participants actually made. This finding rules out the possibility that the Study 1 meta-perception findings were driven by individual differences in meta-perceptions among people who rejected versus accepted harm, thereby overcoming the limitation of employing quasi-experimental designs. However, thus far we have examined meta-perceptions using only the crying baby dilemma in American MTurk samples. To improve generalizability, we examined whether these effects would replicate using a whole battery of dilemmas and an in-lab sample of German-speaking student participants.

4.7 Study 3

Study 3 examined whether the meta-perception findings in Studies 1 and 2 would generalize to other dilemmas and samples. Thus, we recruited a laboratory sample of German-speaking university students and broadened the stimulus set by translating a standardized battery of 10 dilemmas into German, and randomly presenting participants with one of the ten dilemmas from this battery (Conway & Gawronski, 2013).

4.7.1 Method

Participants. We obtained 131 German university students (55 males, 75 females, 1 other, 2 no gender indication, M_{age} = 30.49, *SD* = 9.90) who received $0.25. Again, we aimed for ~50 participants per cell, and excluded no one ($n_{\text{harm_rejection}}$ = 66; $n_{\text{harm\_acceptance}}$ = 65). Again, we had ~99% power to detect the obtained interaction.

Procedure. The design was similar to the meta-perceptions condition in Study 1. Each participant read one dilemma at random from a battery of 10 dilemmas, selected either accept or reject outcome-maximizing harm as in Study 1, and completed the same meta-perception measures of how others would view their warmth (α = .89) and competence (α =.87). This time we did not measure morality. The battery consisted of all 10 incongruent dilemmas from Conway and Gawronski (2013), where causing harm always maximized outcomes. The crying baby and vaccine dilemmas from Study 1 are examples of incongruent dilemmas from this set. Other examples include the *torture dilemma* (is it appropriate to torture a man in order to stop a bomb that will kill people?), and the *car accident dilemma* (is it appropriate to run over a grandmother in order to avoid running over a mother and child?). This design resulted in a 2 (decision: harm rejection vs. acceptance) × 2 (dimension: warmth vs. competence) quasi-experimental design. Again, warmth and competence correlated moderately (*r* = .28).

4.7.2 Results

We submitted warmth and competence ratings to a 2 (decision: harm rejection vs. acceptance) × 2 (dimension: warmth vs. competence) repeated measures ANOVA with the first factor between-subjects and the second factor within-subjects. There was a main effect of target decision: Participants who rejected harm, upholding deontology, reported higher meta-perception ratings overall (*M* = 4.76, *SD* = 1.14), than participants who accepted harm, upholding utilitarianism (*M* = 4.16, *SD* = 1.26), *F*(1, 187) = 6.44, *p* = .012, η_p^2 = .03. There was also main effect of personality dimension: participants reported lower overall meta-perception ratings for warmth (*M* = 3.74, *SD* = 1.76) than competence (*M* = 4.78, *SD* = 1.16), *F*(1, 187) = 9.06, *p* = .003, η_p^2 = .05. More importantly, these results were qualified by the predicted interaction, *F*(1, 187) = 42.58, *p* < .001, η_p^2 = .19.

Post-hoc tests revealed the same warmth/competence tradeoff as in Studies 1 and 2: Participants expected that others would rate them as warmer when they rejected (*M* = 5.02, *SD* = 1.49) rather than accepted causing harm (*M* = 3.48, *SD* = 1.71), *F*(1, 129) = 25.75, *p* < .001, η_p^2 = .17. Conversely, participants expected that others would rate them as less competent when they rejected (*M* = 4.51, *SD* = 1.19) rather than accepted causing harm (*M* = 4.82, *SD* = 1.19), *F*(1, 129) = 13.98, *p* < .001, η_p^2 = .01.

4.7.3 Discussion

Study 3 replicated and generalized the meta-perception findings from Studies 1 and 2 to a different sample and broader array of dilemma stimuli. These findings increase confidence in the claim that participants in both Germany and the United States hold accurate meta-perceptions regarding how their judgments on many dilemmas make them appear to others—namely, participants are aware of the warmth/competence perception tradeoff implied by dilemma judgments. Next, we turn to the possibility that people use this meta-perception information to adjust their dilemma decisions to strategically present themselves as relatively warm or competent depending on which trait is most valued in a given context.

4.8 Study 4

In Study 4, we examined whether people sometimes strategically adjust their private dilemma judgments to mesh with social expectations. We randomly assigned participants to learn that the study was ostensibly comparing either the intellectual or emotional abilities of people in different university degree programs. Part of the study involved responding to moral dilemmas. If people consider self-presentation when answering dilemmas, they should be more likely to reject harm when they think the study assessed emotional competency (i.e., warmth), and more likely to accept outcome-maximizing harm when they think the study is about intellectual ability (i.e., competence).

4.8.1 Method

Participants. We obtained 120 German participants (57 males, 63 females, M_{age} = 22.99, *SD* = 4.41) from a large University in Western Germany, who received € 2.00. Participants were randomly assigned to a condition prioritizing logical reasoning (associated with competence) or emotional competency (associated with warmth, Rom et al., 2017). We again aimed to collect ~50 participants per cell, though we ended up with a few extra people. No participants were excluded (n_{emotion} = 60; n_{logic} = 60). Again, Gpower suggested we had ~99% power to detect the obtained interaction.

Procedure. To manipulate the perceived importance or warmth and competence, we randomly assigned participants to read the following instructions: *This is a study to measure the logical reasoning ability (or emotional competency) between people in different degree programs. Please imagine the following situation and tell us your solution.* Then we presented them with three dilemmas, presented on individual screens, in a fixed random order. We again employed the same ten dilemmas by randomly presenting dilemmas from a standardized battery as in Study 3 (Conway & Gawronski, 2013). Participants indicated how much they accepted vs. rejected such outcome-maximizing harm on scales from *harm is not acceptable* (1) to *harm is acceptable* (7). We averaged ratings to form an aggregate score of relative harm acceptability (α = .55). Although this reliability is lower than ideal, it is to be expected for such a wide range of content and few datapoints, and makes the analysis more conservative.

4.8.2 Results and Discussion

As predicted, participants indicated that causing harm to maximize outcomes was relatively more appropriate in the condition emphasizing logic (*M* = 4.74, *SD* = 1.50), versus emotional ability (*M* = 4.06, *SD* = 1.45), *t*(118) = -2.53, *p* = .013, *d* = 1.70. This finding provides initial evidence suggesting that participants may modify dilemma answers to present themselves as relatively warm or competent, depending on which trait is prioritized in the current context. However, it remains unclear whether this effect reflects strategic self-presentation or whether the instructions simply primed participants to focus more on emotion or logic when forming judgments (similar to Valdesolo & DeSteno, 2007). Therefore, in Study 5 we assessed not only which dilemma decision participants report, but also which decision they believed others expected them to make. If strategic self-presentation plays a role in dilemma judgments, then both actual and expected decisions should reflect the influence of the context manipulation. We also employed a new manipulation to increase generalizability.

4.9 Study 5

In Study 5 we examined whether both expected and reported dilemma judgments conform to social role expectations. Participants imagined they were applying for a job as a military physician, and one interview question involved a moral dilemma. We manipulated whether warmth or competence was valued most by emphasizing either the military (competence) or medical treatment (warmth) aspects of the position. Then participants reported which dilemma answer they thought interviewers expected, as well as their actual decision. If people adjust their dilemma judgments to conform to social role expectations, then participants in the military condition should be more likely to infer and report accepting harm to cause outcome-maximizing harm than in the physician condition.

4.9.1 Method

Participants and design. Again, to increase confidence in the findings of this conceptual replication, we roughly doubled sample size to 200 American Mechanical Turk participants (n_{military} = 100; $n_{\text{physician}}$ = 100), who received $0.25 (128 males, 72 females, M_{age} = 33.23, *SD* = 10.89). This time, we predicted a main effect rather than interaction; Gpower indicates this design again provided ~99% power to detect the predicted main effect. We randomly assigned participants to either the military or physician emphasis conditions. No data were excluded.

Procedure. We asked participants to imagine they were interviewing for a job they really wanted, and gave them one of two job descriptions adapted from past work on masculine and feminine job descriptions (Rudman & Glick, 1999). In the military condition participants read (bold in original): *As a **military** physician you will be responsible for the health and well-being of personnel in your military unit. On the battlefield, soldiers are in harm's way. There will be casualties. The ideal **military** doctor is technically **skilled**, **ambitious, strongly independent**, and able to perform well under pressure.* In the physician condition participants read: *As a military **physician** you will be responsible for the health and well-being of personnel in your military unit. On the battlefield, soldiers are in harm's way. There will be casualties. The ideal military **doctor** is technically skilled and able to work well under pressure, but also **helpful**, **sensitive** to the needs of each individual patient, and able to **listen carefully** to their patients' concerns.*

Next, participants imagined they must answer a moral dilemma as part of the interview process. We presented a version of the transplant dilemma where a surgeon could allow one ill patient to die to use their organs to save five other patients (Greene at el., 2001). Participants reported their perception of interviewer expectations on two scales from 1 (*not at all*) to 7 (*very much*): *How much does the interviewer want you to say **YES (NO)**; that withholding the medical care from Patient 6 in order to save the other five patients is **(NOT)** appropriate?*

We measured expectations twice using different framings in case participants viewed these as independent questions. However, they strongly negatively correlated (*r* = -.79), so we reverse-coded the second question and combined them into a single measure reflecting increased acceptance of outcome-maximizing harm. Finally, participants indicated their actual judgment on the same scale as Study 4.

4.9.2 Results

We conducted a 2 (condition: physician vs. military emphasis) × 2 (decision: expectation vs. answer) repeated measures ANOVA (see Figure 3) with the first factor between-subjects and second factor within-subjects. This analysis yielded the expected main effect of condition: participants gave higher harm acceptance ratings (upholding utilitarianism/rejecting deontology) in the military (*M* = 4.10, *SD* = 2.33) than physician emphasis condition (*M* = 3.09, *SD* = 1.80), *F*(1, 198) = 11.63, *p* = .001, η_p^2 = .06. We also found an unexpected main effect for decision: overall, participants reported lower harm acceptance ratings for expectations (*M* = 3.41, *SD* = 2.03) than their real answers (*M* = 3.80, *SD* = 2.62), *F*(1, 198) = 8.13, *p* = .005, η_p^2 = .04. The emphasis condition × decision type interaction was not significant, *F*(1, 198) = 1.07, *p* = .301, η_p^2 = .01. Post-hoc tests confirmed that participants thought the interviewers expected them to accept harm more in the military (*M* = 3.84, *SD* = 2.19) than physician emphasis condition (*M* = 3.00, *SD* = 1.76), *F*(1, 198) = 9.49, *p* = .002, η_p^2 = .05; likewise, participants were more likely to actually accept harm in the military (*M* = 4.36, *SD* = 2.77) than physician emphasis condition (*M* = 3.21, *SD* = 2.34), *F*(1, 198) = 10.01, *p* = .002, η_p^2 = .05.

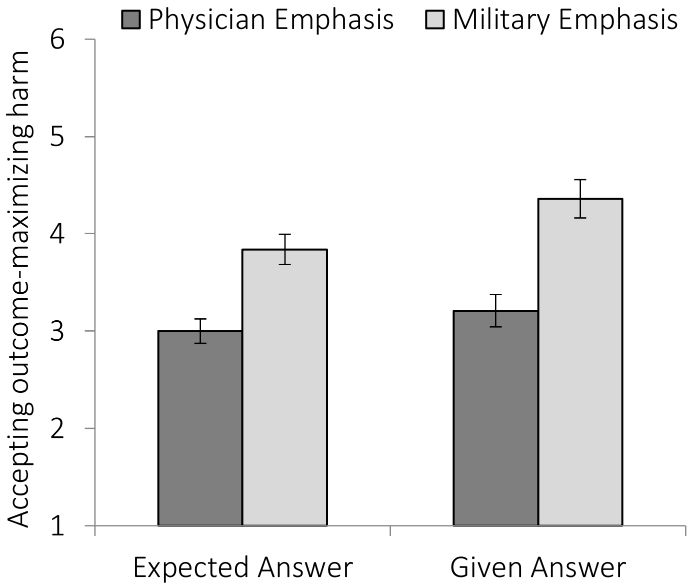


Figure 3. Mean expected and actual dilemma decisions (accepting outcome-maximizing harm, upholding utilitarianism) during military physician job interview emphasizing either military or physician skills, Study 5. Error bars reflect standard errors.

4.9.3 Discussion

When a military physician job description emphasized sensitive caring, participants expected and indicated that causing harm to maximize outcomes was less acceptable to job interviews; when a description of the same job emphasized ambitious independent skill, participants expected and indicated that causing harm to maximize outcomes was more acceptable to job interviews. These findings replicate Study 4 using a different manipulation, providing further support for the argument that participants modify dilemma judgments to present themselves favorably. However, it remains possible that features of the job description simply primed participants to consider emotion or logic more carefully when forming both expectation judgments and actual dilemma judgments. In order to demonstrate *strategic* self-presentation, it is necessary to demonstrate that whereas participants' public dilemma decisions accord with expectations, their private decisions do not. We examined this possibility in Study 6.

4.10 Study 6

Study 6 employed a similar design to Study 5, except for two changes. This time we assessed private dilemma decisions in addition to public dilemma decisions and perceived expectations. Second, we employed yet another method of manipulating whether warmth or competence traits were situationally valued: participants imagined applying for a prestigious scholarship that emphasized either warmth or competence.[22] If people employ strategic self-presentation when forming dilemma judgments, participants who learn the scholarship foundation values warmth should both expect and publicly select harm rejection judgments more often, whereas participants who learn the foundation values competence should both expect and publicly select harm acceptance judgments more often—however, private judgments should remain unaffected. Conversely, if the manipulation simply primed participants to differentially consider emotion or logic when forming their answer, then private judgments should evince the same pattern as expectations and public judgments.

4.10.1 Method

Participants. We obtained 200 American (117 males, 83 females, M_{age} = 32.42, SD = 11.30) participants via Mechanical Turk, who received $0.25. Again, this design provided ~99% power to detect the obtained interaction. Participants were randomly assigned to one of two conditions: warmth or competence emphasis ($n_{\text{warmth_emphasis}}$ = 100; $n_{\text{competence\_emphasis}}$ = 100).

Procedure. Participants imagined they were interviewing for a prestigious scholarship they *really wanted*. They read: *You are interviewing with the National Merit Foundation for a prestigious fellowship. It is extremely important for you to get the fellowship and you have been training a long time to get it. During the interview, you remember what kind of person they are looking for*. The competence emphasis condition continued, *"The ideal scholar of our*

[22] This manipulation was derived from a real life experience of the first author: she had to complete a moral dilemma while applying for a prestigious fellowship, and tried to guess which answer was expected.

foundation is ***skilled****,* ***ambitious****, and a good* ***thinker****,"* whereas the warmth emphasis condition continued: *"The ideal scholar of this foundation is* ***good-natured****,* ***helpful****, and a good* ***listener****"* (bold in original).

Participants then imagined they had answered numerous questions about their background and the interview was going well—but one final question remained that would impact whether or not they would obtain the scholarship: the vaccine dilemma employed previously. Participants read that dilemma, then indicated a) which answer they thought the interviewers *expected*, b) which answer they would *privately make*, and c) which answer they would *publicly make* in front of the interviewers. Participants indicated each answer on 7-point scales from *harm is not appropriate* (1) to *harm is appropriate* (7).

4.10.2 Results

We conducted a 2 (emphasis: warmth vs. competence) × 3 (decision type: expectation vs. private judgment vs. public judgment) repeated measures ANOVA with the first factor between and the second factor within participants (see Figure 4). This analysis yielded a main effect of condition: participants gave lower overall harm acceptability ratings in the warmth (M = 5.40, SD = 1.91) than competence emphasis condition (M = 4.75, SD = 1.92), F(2, 197) = 3.21, p = .042, η_p^2 = .03. There was no main effect for decision type, F(2, 197) = 2.64, p = .072, η_p^2 = .013. However, results were qualified by the expected interaction, F(2, 197) = 6.65, p < .001, η_p^2 = .05.

Post-hoc comparisons indicated that both expectations and public judgments replicated Study 5: participants in the warmth emphasis condition reported that interviewers expected less harm acceptance (M = 4.60, SD = 2.31), than participants in the competence emphasis condition (M = 5.64, SD = 2.24), F(1, 198) = 10.29, p =.002, η_p^2 = .05. Likewise, participants in the warmth emphasis condition were less likely to publicly indicate acceptance of outcome-maximizing harm (M = 4.37, SD = 2.26), than participants in the competence emphasis condition (M = 5.40, SD = 2.36), F(1, 198) = 10.10, p = .002, η_p^2 = .05. However, private judgments remained unaffected by the manipulation: participants in the warmth emphasis condition (*M* = 5.28, *SD* = 2.25) did not significantly differ from participants in the competence emphasis condition regarding harm acceptability (*M* = 5.16, *SD* = 2.28), *F*(1, 198) = .13, *p* = .719, η_p^2 = .00.

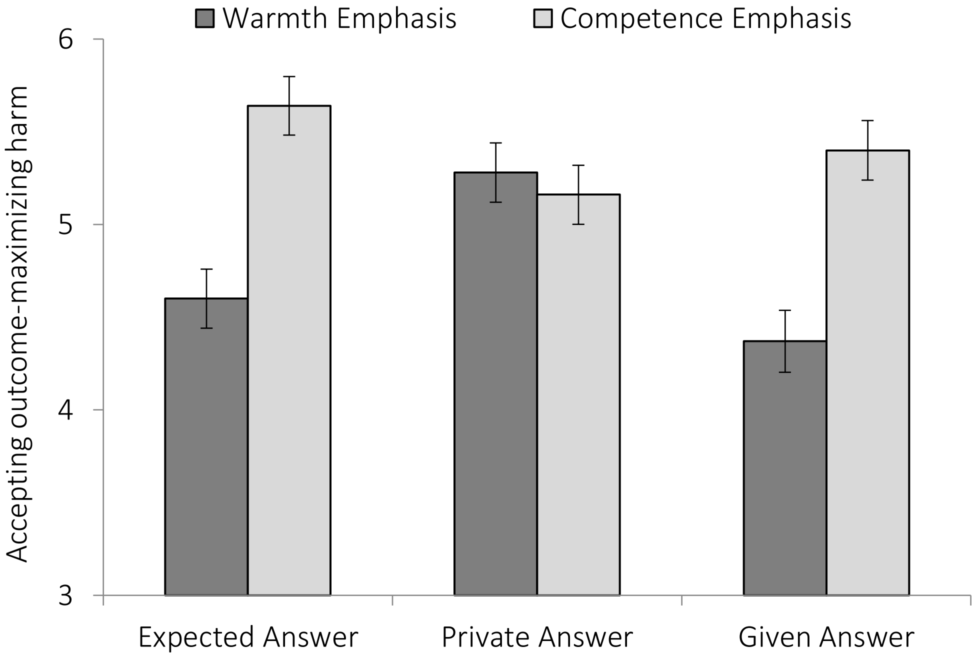


Figure 4. Mean expected, private, and public dilemma decisions (accepting outcome-maximizing harm, upholding utilitarianism) when applying for a scholarship emphasizing either warmth or competence, Study 6. Error bars reflect standard errors.

4.10.3 Discussion

Study 6 replicated and extended the findings of Studies 4 and 5, providing increased support for our argument that participants strategically adjust dilemmas judgments in order to present situationally favorable impressions. Using a different manipulation, we again found that participants both expected and publicly made fewer harm-acceptance judgments when the situation emphasized warmth than competence. However, private judgments remained unaffected by the manipulation. These findings rule out the possibility that the differences in expectation and public judgment reflect priming, as private judgments remained unaffected by the manipulation. Instead, these findings suggest that participants employed strategic self-presentation to publicly provide dilemma answers that accorded with expectations rather than private considerations.

4.11 Study 7

Together with past findings (Rom et al., 2017), the current work suggests that dilemma answers entail a warmth/competence trade-off: rejecting harm makes one appear warm but less competent, whereas accepting outcome-maximizing harm makes one appear cold but more competent. Although people appear to strategically modify their dilemma judgments to emphasize either warmth or competence, doing so entails the trade-off of appearing weaker on the converse trait. Yet, on many occasions it may be optimal to present oneself as high on both traits. For example, politicians may wish to appear both warm and competent to increase chances of re-election, yet face dilemmas that pit individual well-being against public interest, such as authorizing forceful interrogations to obtain life-saving information, or directing medical funding away from rare but deadly disorders towards widespread problems. Is it possible to frame one's dilemma decision so as to reduce the warmth/competence trade-off previously obtained? Everett and colleagues (2016) provided initial evidence consistent with this possibility: when an injured soldier begged to death to avoid capture and torture by the enemy, and decision-makers rejected this request, perceivers viewed them as more moral and trustworthy when they offered categorical deontological justifications (i.e., "killing is wrong even if it has good consequences") compared to utilitarian or contractual reasons. These arguments do not speak directly to perceptions of warmth or competence, but they suggest that perceivers draw inferences from the justifications decision-makers provide, beyond the decisions they make.

We hypothesized that decision-makers can augment perception of their weaker trait by supplementing their dilemma decision itself with justifications that appeal to emotions or to logic. If such appeals impact dilemma perceptions, then people who accept causing outcome-maximizing harm may appear less cold by expressing emotional concern for the victim of harm compared to people who accept causing harm without expressing emotions. Conversely, people who reject harm may appear less incompetent by expressing consideration of logical reasoning compared to people who reject harm without expressing logical reasoning. Hence, we predicted a three-way interaction between dilemma decision, justification, and trait measure. To examine this possibility, we assessed warmth and competence perceptions of dilemma decision-makers who either accepted or rejected harm, and who framed their decision either in terms of emotion or cognition.

4.11.1 Method

Participants and design. We obtained 401 American (251 males, 150 females, M_{age} = 32.42, *SD* = 11.30) participants via Mechanical Turk, who received $0.30. Participants were randomly assigned to learn that a previous participant either accepted or rejected harm for either emotional or logical reasons, and rated them on warmth and competence, for a 2 (explanation: emotional vs. logical) × 2 (target decision: harm rejection vs. acceptance) × 2 (personality dimension: warmth vs. competence) design, where the first two factors varied between-subjects and the final factor varied within-subjects. Despite the large sample, GPower indicated that this study had only ~82% power to detect the obtained three-way interaction ($n_{\text{emotional}}$ = 100; n_{logical} = 101).

Procedure. The procedure was similar to the other-perception condition in Study 1: Participants viewed a photo of a university student named Brad who ostensibly previously participated. They read the crying baby dilemma and learned that Brad ostensibly either accepted or rejected the specified harm, accompanied by a brief written explanation emphasizing either emotional or logical justifications for this decision.

Specifically, in the emotional harm rejection condition, participants read, "*No, it is completely unacceptable to kill the baby! It doesn't matter what the reasons are; I just could not live with myself if I hurt an innocent little baby. Killing is forbidden for any reason and never justified.*" In the in logical harm rejection condition participants read, "*No, it is unacceptable to kill the baby! I understand that doing so makes logical sense, but killing some people to protect others creates an immoral society. It is better to live in a society that forbids killing for any reason than one where killing some people is justified to help others*." In the emotional harm acceptance condition participants read, "*Yes, it is acceptable to kill the baby. It is true that it would break my heart to kill an innocent baby, but it just makes sense to perform the action that saves everybody. It upsets me very much, but it's the only logical thing to do.*" Finally, in the rational harm acceptance condition participants read, "*Yes, it is completely acceptable to kill the baby! It just makes sense to perform the action that saves everybody. It's the only logical thing to do.*" After reading Brad's dilemma decision and justification, participants rated Brad's warmth (α = .87) and competence (α = .82) as before.

4.11.2 Results

Target warmth and competence. We submitted ratings to a 2 (justification type: emotional vs. logical) × 2 (target decision: harm rejection vs. acceptance) × 2 (personality dimension: warmth vs. competence) repeated-measures ANOVA with the first two factors between-subjects and the last factor within-subjects (see Figure 5). There was no main effect for justification type, $F(1, 397) = 0.00$, $p = .990$, $\eta_p^2 = .00$, or personality dimension, $F(1, 397) = .86$, $p = .355$, $\eta_p^2 = .00$. However, there was a main effect of target decision: participants gave higher ratings overall when Brad rejected ($M = 5.10$, $SD = 1.10$) versus accepted causing harm ($M = 4.70$, $SD = .99$), $F(1, 397) = 15.13$, $p < .001$, $\eta_p^2 = .04$. These results were qualified by significant two-way interactions between justification type and personality dimension, *F*(1, 397) = 31.91, $p < .001$, $\eta_p^2 = .07$, and between target decision and personality dimension, *F*(1, 397) = 260.04, $p < .001$, $\eta_p^2 = .40$. The three-way interaction did not approach conventional levels of significance, *F*(1, 397) = 2.27, $p = .136$, $\eta_p^2 = .01$, so we examined the two-way interactions.

The first interaction indicated that justifications impacted warmth and competence decisions: perceivers rated Brad as higher on warmth when he provided emotional justifications (*M* = 5.05, *SD* = 1.30) than when he provided rational justifications (*M* = 4.78, *SD* = 1.34), *F*(1, 399) = 4.28, $p = .039$, $\eta_p^2 = .01$, whereas they rated him higher on competence when he provided rational (*M* = 5.01, *SD* = 1.06) than emotional justifications (*M* = 4.74, *SD* = 1.21), *F*(1, 399) = 5.61, $p = .018$, $\eta_p^2 = .01$.

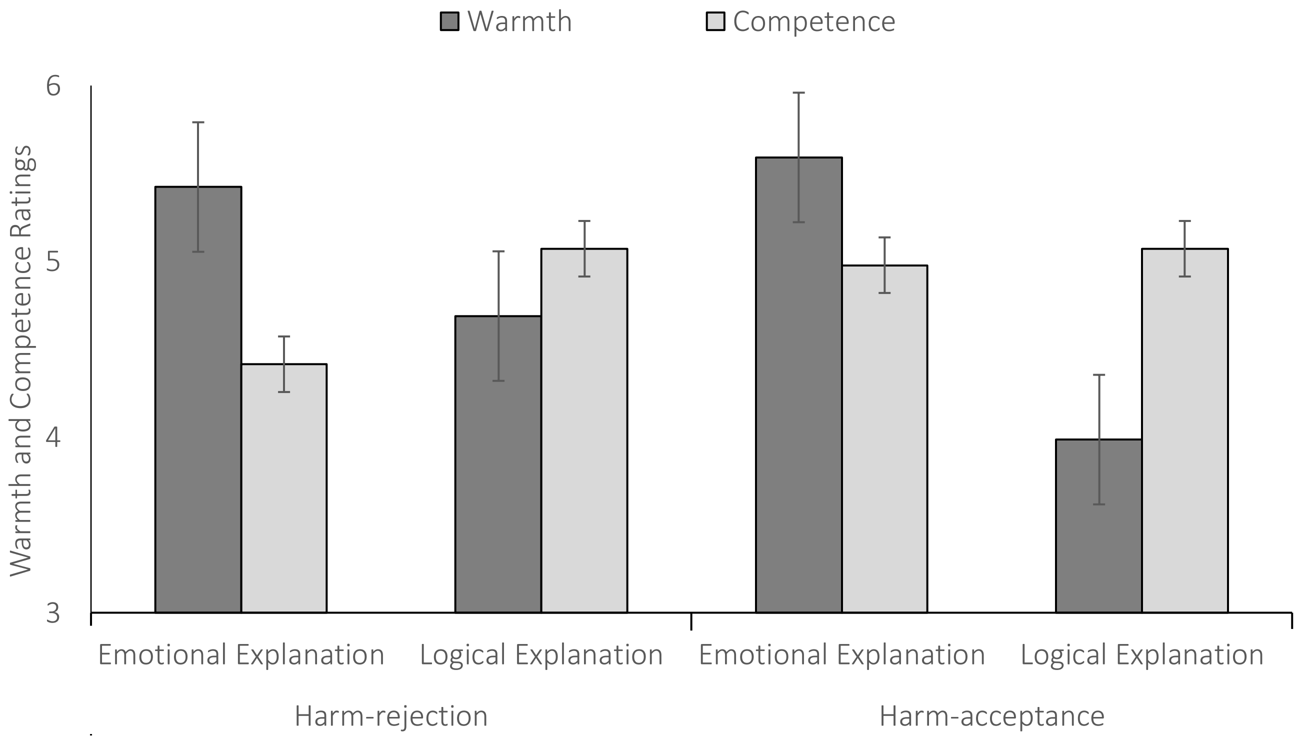


Figure 5. Warmth and competence ratings when Brad rejected or accepted causing harm and either gave an emotional or logical explanation, Study 7. Error bars reflect standard errors.

The second interaction replicated previous findings by showing that dilemma decisions impacted warmth and competence ratings: perceivers rated Brad as higher on warmth when he rejected outcome-maximizing harm (upholding deontology) (M = 5.51, SD = 1.15) than when he accepted outcome-maximizing harm (upholding utilitarianism) (M = 4.34, SD = 1.24), F(1, 399) = 96.16, p < .001, η_p^2 = .19, whereas they rated him as higher on competence when he accepted outcome-maximizing harm (upholding utilitarianism) (M = 5.10, SD = .98) than when he rejected outcome-maximizing harm (upholding deontology) (M = 4.69, SD = 1.27), F(1, 399) = 10.47, p = .001, η_p^2 = .03.

4.11.3 Discussion

Our hypothesis was partially supported. Although we did not find the anticipated predicted three-way interaction, the two-way interaction between justification and trait measure suggested that emotional justifications boosted perceptions of warmth, and logical justifications boosted perceptions of competence, regardless of target decision. Moreover, the two-way interaction between dilemma decision and trait measure replicated the warmth/competence trade-off implied by dilemma decisions demonstrated in previous work (Rom et al., 2017). These findings suggest that decision-makers may be able to bolster perceptions of warmth or competence via relevant justifications, independent of their actual judgment. Hence, decision-makers who accept harm may be able to offset concerns about their warmth by framing their decision in emotional terms, and decision-makers who reject harm may be offset concerns about their competence by framing their decision in logical terms. These findings align with those of Everett and colleagues (2016), suggesting that beyond the judgment itself, perceivers care about *why* decision-makers arrived at their judgment.

4.12 General Discussion

Across seven studies, we garnered evidence that people hold accurate meta-perceptions regarding whether their dilemma decisions convey warmth or competence, and strategically adjust dilemma judgments to present themselves favorably. Study 1 replicated past work (Rom et al., 2017) by demonstrating participants typically rated decision-makers who rejected harm (upholding deontology) as warmer and more moral, but less competent, than decision-makers who accepted outcome-maximizing harm (upholding utilitarianism), together with the novel finding that participants anticipated that others would rate them according to the same warmth/competence tradeoff following the same respective decisions—even though, privately, participants rated themselves as high on both warmth and competence regardless of their decision. Moreover, participants anticipated that others' warmth and competence ratings would reflect whichever judgment those others learned participants made, even when this belief was erroneous (Study 2). Importantly, meta-perceptions of this warmth/competence trade-off generalized to a battery of various dilemma stimuli and a different sample (Study 3). Thus, it seems clear that people hold accurate meta-perceptions regarding how others perceive them based on their dilemma judgments, that these meta-perceptions differ from self-perceptions, track information available to others, and do not merely reflect individual differences in which judgments people prefer.

Next, we examined whether people use meta-perception information to strategically adjust their judgments. First, we demonstrated that dilemma decisions are sensitive to context: When we framed the Study 4 as focusing differences in emotional competency, participants were more likely to reject causing harm (upholding deontology), thereby emphasizing their warmth and emotional processing, compared to when the study was framed as examining differences in logical reasoning, when participants were more likely to accept harm (upholding utilitarianism), thereby emphasizing their competence and logical skills. Study 5 replicated this finding using a different manipulation, where participants simulated interviewing for a job as a military physician, where the description emphasized either military competency or physician care. Participants were more likely to reject harm in the care than competency condition. Study 6 replicated both of these effects using yet another manipulation—a scholarship application that emphasized either academic competency or interpersonal skills. Importantly, this manipulation influenced both expectations and public judgments—but failed to impact private judgments, suggesting that participants were *strategically* adjusting dilemma judgments rather than merely responding to primes in the stimulus materials.

Finally, Study 7 demonstrated that decision-makers can use communication strategies to augment relevant trait. Specifically, decision-makers can provide either emotional or logical justifications for either dilemma judgment, and these justifications impact perceptions of warmth and competence independent of the decision they make. Hence, decision-makers who accept harm (upholding deontology) can offset perceptions of incompetence by describing logical reasons for their decision, whereas decision-makers who accept outcome-maximizing harm (upholding utilitarianism) can offset perceptions of coldness by describing emotional experiences.

In each case, the impacts of dilemma decisions on warmth perceptions was mirrored by similar patterns on ratings of decision-maker *morality*. Indeed, warmth and morality correlated highly in all studies. Such findings could be taken as evidence that warmth and morality reflect a single core construct, but recent work suggests that lay people draw important distinctions between warmth/sociability (i.e., interpersonal friendliness) and morality (e.g., trustworthiness—see Brambilla et al., 2011; Goodwin et al., 2014). Such findings could also reflect the possibility that the current measure of warmth actually reflects moral character evaluations instead of genuine perceptions of warmth/sociability, by including items such as *sincere*. However, as noted above, re-analyses employing revised composites excluding such terms, or indeed using only the single item *warm* demonstrate the same pattern as the warmth composite using all warmth items. Therefore, we suggest that perceivers draw inferences of both warmth and morality from others' deontological dilemma judgments, and these inferences happen to covary substantially in the current paradigm. It may be that these inferences stem from different aspects of deontological judgments—perhaps warmth perceptions reflect inferences of emotional processing, whereas morality inferences reflect perceptions of rule-following—which covary in the current paradigm. Consistent with this possibility, Rom and colleagues (2016) found that perceptions of emotional processing mediated the effect of dilemma judgment on perceptions of warmth—but not on perceptions of morality. Future work might profit from disentangling which aspects of deontological decision-making imply warmth and which imply morality.

4.12.1 Implications for Models of Moral Judgment

The dual-process model of moral judgment (Greene et al., 2001) and other popular models (Cushman, 2013; Crockett, 2013; Sunstein, 2005; Mikhail, 2007) describe the impact of basic psychological processes on moral dilemma judgments, such as affective reactions to harm, cognitive evaluations of outcomes, or heuristic application of moral rules. Importantly, all of these processes should apply similarly whether participants respond to moral dilemmas alone on a desert island or during a live television broadcast watched by millions. We do not dispute the importance of basic processes for influencing dilemma judgments, but we suggest that existing theories are incomplete if they treat public versus private circumstances as identical. We suggest that answering dilemmas while on television—or in any social situation—evokes concern over others' perceptions of ones' dilemma judgment, and how that judgment reflects on oneself. People appear aware of the warmth/competence trade-off others infer from their decision, and strategically modify judgments to present themselves favorably. Hence, higher-order social processes causally contribute to dilemma responses, in addition to basic processes.

The finding that strategic self-presentation drives variance in dilemma judgments suggests that researchers should revisit earlier findings to consider whether self-presentation may account for some of the variance ascribed to basic processes. For example, various researchers have documented gender differences in dilemma judgments (e.g., Fumagalli et al., 2010; Arutyunova, Alexandrov, & Hauser, 2016), such that women evince stronger inclinations to reject harm than men, but similar inclinations to maximize outcomes, leading to higher reports of conflict (Friesdorf, Conway, & Gawronski, 2015). Typically, researchers explain such gender differences in terms of biologically-based constructs such as empathy (Eisenberg & Lennon, 1983) and testosterone (Carney & Mason, 2010), or differences in socialization practices (Eagly & Wood, 1999). However, the current findings raise an alternative possibility: it may be that women experience stronger social expectations to avoid causing harm than do men, even as they appreciate the logic of doing so. After all, women often face pressure to appear both warm and competent, whereas often competence alone often meets male role expectations (Rudman & Glick, 1999). Moreover, women often feel more obliged to engage in self-presentation than do men (Deaux *&* Major, 1987). Such expectations could lead women to reject harm (upholding deontology) more frequently, despite experiencing similar basic processing as do men.

In other work, Lucas & Livingstone (2014) found that participants who socially connected with others made more utilitarian judgments, presumably because social connection reduced aversive affect associated with deontological judgments. However, our results suggest an alternative process: participants who had already connected with others may have felt they established sufficient evidence of warmth or morality that they could afford to display other qualities, such as competence. Indeed, such alternative explanations may occur even in studies where there is no direct social contact between participants and others (e.g., online studies). From a Griceian (1989) perspective, every research study is effectively an act of social communication between the participant and the experimenter. Cues in the framing, instructions, or manipulations of any dilemma study may hint at whether warmth or competence is contextually prioritized, leading participants to infer that one or another dilemma answer is preferred.

Indeed, self-presentation of this sort may even account for some of the response variance between the original trolley dilemma, where approximately 80% of people accept causing harm to save lives, and the footbridge dilemma, where about 80% of people reject causing harm (Greene et al., 2001). In the footbridge dilemma, harm acceptance means being willing to push and thereby kill with one's own hands, whereas in the trolley dilemma harm, acceptance means simply pressing a button. Research suggests that employing the personal force of one's physical being to kill another is more aversive than employing a mechanical mediator (Greene et al., 2009). Accordingly, lay perceivers may view harm caused through personal force as more likely evidence of cold-heartedness than harm caused through intermediaries—thereby creating greater social pressure to avoid causing harm on the footbridge than trolley dilemma. Consistent with this possibility, Everett and colleagues (2016) found that perceivers drew important distinctions between the trustworthiness of decision-makers who accepted vs. rejected harm on the footbridge dilemma, but less of a distinction between those who accepted vs. rejected harm on the trolley dilemma. Future work should directly examine social expectations of appropriate answers in such cases.

4.12.2 Limitations

This research shares limitations with nearly all dilemma research: of necessity, participants make decisions about hypothetical scenarios rather than actual situations. Hence, it remains possible that perceptions and meta-perceptions of real-life dilemma decisions (such as Turing's decision from the beginning of the paper) evince different or even stronger effects. In addition, like most dilemma research, the dilemmas employed here vary on a number of factors that may influence judgments, such as whether the victim of harm is guilty of causing danger or not, or is fated to die or not (Christensen, Flexas, Calabrese, Gut, & Gomila, 2014). Future work should systematically vary each of these factors to determine whether they impact perceptions and meta-perceptions of dilemma judgments. Moreover, the dilemmas employed here examine only violations of moral proscriptions—causing harm to maximize outcomes—whereas it is possible to conceptualize dilemmas involving prescription violations—saving one person at a risk to many—that may entail different perceptions and meta-perceptions (Gawronski et al., 2015). Future work may profit by comparing such dilemmas.

In addition, although the current work employed participants from different countries in several languages, all participants hailed from broader 'Western' culture. Recent work has documented that East Asian participants are less likely to endorse outcome-maximizing harm than Western participants (e.g., Gold, Colman, & Pulford, 2014). One reason for this difference may be increased fatalism in Asian culture—the belief that one should not interfere with destiny (Chih-Long, 2013). It remains unclear whether perceptions of dilemma judgments also reflect such cultural variation—the cultural background of both perceivers and decision-makers may matter. Future research might profitably investigate these possibilities.

4.12.3 Conclusion

Building on work examining the role of basic psychological processes in driving dilemma judgments, the current work provides evidence that higher-order social processes also play a role. Participants demonstrated accurate meta-insight into how warm and competent their dilemma judgments would make them appear to others, and strategically shifted public (but not private) dilemma judgments to accord with such expectations depending on whether situations prioritized warmth or competence. These findings suggest that classic models of dilemma decision-making (e.g., Greene et al., 2001) underestimate the influence of social considerations. When Allied forces allowed the Axis raid on Coventry to proceed so as to protect the Enigma Code decryption, they likely engaged in not only basic emotional and logical processing, but also considered how their allies would have reacted to this decision. In the midst of a desperate war, they selected a decision that made them appear competent at the cost of warmth—had circumstances been different, perhaps they would have selected an entirely different answer.

4.13 References

Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science*, *23*, 861-868. doi:10.1177/0956797611434965

Anderson, C., Ames, D. R., & Gosling, S. D. (2008). Punishing hubris: The perils of overestimating one's status in a group. *Personality and Social Psychology Bulletin, 34*, 90-101. doi: 10.1177/0146167207307489

Andersen, S. M., & Ross, L. (1984). Self-knowledge and social inference: I. The impact of cognitive/ affective and behavioral data. *Journal of Personality and Social Psychology*, *46,* 280-293.

Asch, S. E. (1948). The doctrine of suggestion, prestige and imitation in social psychology. *Psychological Review*, 55, 250–276.

Aquino, K., & Reed, A. II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*, 1423-1440. doi:10.1037//0022-3514.83.6.1423

Arutyunova, K. R., Alexandrov, Y. I., & Hauser, M. D. (2016). Sociocultural Influences on Moral Judgments: East–West, Male–Female, and Young–Old. *Frontiers in Psychology*.

Barish, K., (Producer), & Pakula, A. J. (Director). (1982). *Sophie's Choice* [Motion picture]. United States: Incorporated Television Company.

Bartels, D. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition, 108*, 381–417. doi:10.1016/j.cognition.2008.03.001

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*, 154-161. doi:10.1016/j.cognition.2011.05.010

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology, 41*, 135-143. Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology, 41*, 135-143.

Bloom, P. (2011). Family, community, trolley problems, and the crisis in moral psychology. *The Yale Review*, *99*(2), 26-43.

Carlson, E. N., & Furr, R. M. (2009). Evidence of differential meta-accuracy: People understand the different impressions they make. *Psychological Science, 20*, 1033-1039.

doi: 10.1111/j.1467-9280.2009.02409.x

Carlson, E. N., Vazire, S., & Furr, R. M. (2011). Meta-insight: Do people really know how others see them? *Journal of Personality and Social Psychology, 101*, 831-846.

doi: 10.1037/a0024297

Carney, D. R., & Mason, M. F. (2010). Decision making and testosterone: When the ends justify the means. *Journal of Experimental Social Psychology*, *46*(4), 668–671. http://doi.org/10.1016/j.jesp.2010.02.003

Chambers, J. R., Epley, N., Savitsky, K., & Windschitl, P. D. (2008). Knowing too much: Using private knowledge to predict how one is viewed by others. *Psychological Science, 19,* 542-548.

doi: 10.1111/j.1467-9280.2008.02121.x

Chih-Long, Y. (2013). It is our destiny to die: The effects of mortality salience and culture-priming on fatalism and karma belief. *International Journal of Psychology*, *48*, 818-828.

doi: 10.1080/00207594.2012.678363

Christensen, J. F., Flexas, A., Calabrese, M., Gut, N. K., & Gomila, A. (2014). Moral judgment reloaded: a moral dilemma validation study. *Frontiers in Psychology, 5*, 1-18. doi:10.3389/fpsyg.2014.00607

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*, 216-235. doi:10.1037/a0031021

Copeland, B. J. (2014). *Turing: pioneer of the information age*. Oxford University Press.

Crockett, M. J. (2013). Models of morality. Trends in cognitive sciences, 17, 363-366.

Cushman, F. (2013). Action, outcome, and value a dual-system framework for morality. *Personality and social psychology review, 17*, 273-292.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*, 1082–1089.

doi: 10.1111/j.1467-9280.2006.01834.x

Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological review*, *94*, 369.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, *109*(3), 573.

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*.

doi:10.1037//0003-066x.54.6.408

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin, 94*, 100–131.

Epley, N., & Dunning, D. (2000). Feeling “Holier than thou”: Are self-serving assessments produced by errors in self or social prediction? *Journal of Personality and Social Psychology*, *79,* 861-875. doi: 10.1037/0022-3514.79.6.861

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87,* 327-339.

doi: 10.1037/0022-3514.87.3.327

Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General, 145*, 772.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2006). Universal dimensions of social cognition: Warmth and Competence. *Trends in Cognitive Sciences, 11*, 77-83. doi:10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878-902.

doi: 10.1037/0022-3514.82.6.878

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5,* 5-15. doi: 10.1093/0199252866.003.0002

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender Differences in Responses to Moral Dilemmas A Process Dissociation Analysis. *Personality and Social Psychology Bulletin*, 0146167215575731.

Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Sposta, S., Zago, S., … Priori, A. (2010). Gender-related differences in moral judgments. *Cognitive Processing*, *11*, 219–226. doi:10.1007/s10339-009-0335-2

Faul, F., Erdfelder, E., Lang, A.-G., &; Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. doi:10.3758/BF03193146

Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R., & Hütter, M. (2015). Moral dilemma judgments: Disentangling deontological inclinations, utilitarian inclinations, and general action tendencies. In J. P. Forgas, P. A. M. Van Lange, & L. Jussim (Eds.), *Social psychology of morality*. New York: Psychology Press.

Gold, N., Colman, A. M., & Pulford, B. D. (2014). Cultural differences in responses to real-life and hypothetical trolley problems. Judgment and Decision Making, 9, 65-76.

Gold, N., Pulford, B. D., & Colman, A. M. (2015). Do as I say, don’t do as I do: Differences in moral judgments do not translate into differences in decisions in real-life trolley problems. Journal of economic psychology, 47, 50-61.

Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. PLOS ONE, 8, 1-9. doi:10.1371/journal.pone.0060418

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*, 148.

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*, 364-371. doi:10.1016/j.cognition.2009.02.001

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389-400. doi: 10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105-2108. doi: 10.1126/science.1062872

Grice, H. P. (1989). Studies in the Way of Words. Cambridge, MA: Harvard University Press.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834. doi:10.1037//0033-295X.108.4.814

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science, 345*(6202), 1340-1343.

Kahane, G. (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social neuroscience*, *10*(5), 551-560.

Kahane, G., Everett, J. A. C., Earp, B. D., Farias, M., &amp; Savulescu, J. (2015). ‘Utilitarian’ judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition, 134*, 193-209. doi:10.1016/j.cognition.2014.10.005

Kant, I. (1785/1959). *Foundation of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis: Bobbs-Merrill.

Kaplan, S. A., Santuzzi, A. M., & Ruscher, J. B. (2009). Elaborative metaperceptions in outcome-dependent situations: The diluted relationship between default self-perceptions and metaperceptions. *Social Cognition, 27*, 601-614. doi: 10.1521/soco.2009.27.4.601

Kenny, D. A., & DePaulo, B. M. (1993). Do people know how others view them? An empirical and theoretical account. *Psychological Bulletin, 114,* 145-161.

doi: 10.1037/0033-2909.114.1.145

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature 446*, 908-911. doi:10.1038/nature05631

Kohlberg, L. (1969). Stage and sequence: The cognitive–developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. (pp. 347–480). Chicago, IL: Rand McNally.

Krebs, D. L. (2011). The evolution of a sense of morality. *Creating consilience*, 299-317.

Kreps, T. A., & Monin, B. (2014). Core values versus common sense consequentialist views appear less rooted in morality. *Personality and Social Psychology Bulletin*, 0146167214551154.

Kruger, J., & Gilovich, T. (2004). Actions, intentions, and self-assessment: The road to self-enhancement is paved with good intentions. *Personality and Social Psychology Bulletin*, *30*(3), 328-339.

Kundu, P., & Cummins, D. D. (2012). Morality and conformity: The Asch paradigm applied to moral decisions. *Social Influence*, *8*, 268-279.

Leary, M. R. (1989). Self-presentational processes in leadership emergence and effectiveness.

Leary, M. R. (1995). *Self-presentation: Impression management and interpersonal behavior*. Brown & Benchmark Publishers.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological bulletin*, *107*(1), 34.

Liu, B. S., & Ditto, P. H. (2013). What dilemma? Moral evaluation shapes factual belief. *Social Psychological and Personality Science*, *4*, 316–323. http://dx.doi.org/10.1177/1948550612456045.

Lucas, B. J., & Galinsky, A. D. (2015). Is utilitarianism risky? How the same antecedents and mechanism produce both utilitarian and risky choices. *Perspectives on Psychological Science*, *10*(4), 541-548.

Lucas, J. L., & Livingstone, R. W. (2014). Feeling socially connected increases utilitarian choices in moral dilemmas. *Journal of Experimental Social Psychology, 53*, 1–4.

doi: 10.1016/j.jesp.2014.01.011

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in cognitive sciences*, *11*(4), 143-152.

Milgram, S. (1963). Behavioral Study of obedience. *The Journal of abnormal and social psychology*, *67*(4), 371.

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *TRENDS in Cognitive Sciences, 11*, 143-152. doi:10.1016/j.tics.2006.12.007

Mill, J. S. (1861/1998). *Utilitarianism*. In R. Crisp (Ed.), New York: Oxford University Press.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science, 19*, 549-57. doi:10.1111/j.1467-9280.2008.02122.x

Peeters, G. (1983). Relational and informational pattern in social cognition. In W. Doise & S. Moscovici (Eds.), *Current issues in European social psychology* (pp. 201-237). Cambridge, England: Cambridge University Press.”.

Pronin, E. (2008). How we see ourselves and how we see others. *Science, 320*, 1177-1180. doi: 10.1126/science.1154199

Reis, H. T., & Gruzen, J. (1976). On mediating equity, equality, and self-interest: The role of self-presentation in social exchange. *Journal of Experimental Social Psychology*, *12*(5), 487-503.

Rom, S. C., Weiss, A., & Conway, P. (2017). Judging those who judge: Perceivers infer the roles of affect and cognition underpinning others' moral dilemma responses. *Journal of Experimental Social Psychology*, *69*, 44-58.

Sarbin, T. R., & Allen, V. L. (1968). Role theory. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (pp. 488–567). Reading, MA: Addison-Wesley.

Sherif, M. A. (1935). A study of some social factors in perception, *Archives of Psychology*, 27, 1–60.

Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making*, *4*(6), 479.

Valdesolo, P. & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17*, 476–477. doi:10.1111/j.1467-9280.2006.01731.x

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*, 281-300. doi: 10.1037/a0017908

Von Baeyer, C. L., Sherk, D. L., & Zanna, M. P. (1981). Impression management in the job interview: When the female applicant meets the male (chauvinist) interviewer. *Personality and Social Psychology Bulletin*, *7*(1), 45-51.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*, 395-412. doi: 10.1037/0022-3514.37.3.395

Winkielman, P., & Schwarz, N. (2001). How pleasant was your childhood? Beliefs about memory shape inferences from experienced difficulty of recall. *Psychological Science*, *12*(2), 176-179.

Winterbotham, F. W. (1974). The Ultra Secret (New York, 1974). *Ronald Lewin, Ultra Goes to War (New York, 1978).*

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. *Advances in experimental social psychology*, *46*, 55-123.

4.14 Supplementary Material

Table S1.

Structure matrix loadings for individual items on the warmth and competence factors derived from principle components analysis with oblimin rotation in all studies using the Fiske et al. (2006) items. Note that Study 3 did not include the item *moral*.

	Tolerant	Warm	Sincere	Good	Competent	Confident	Competitive	Independent	Intelligent	Moral
Study 1										
Warmth	**.778**	**.909**	**.870**	**.931**	.361	.216	.137	.273	.298	**.842**
Competence	.303	.239	.368	.249	**.842**	**.832**	**.648**	**.839**	**.829**	.205
Study 2										
Warmth	**.913**	**.939**	**.937**	**.950**	**.697**	**.695**	.206	.221	**.654**	**.949**
Competence	.388	.299	.477	.365	**.781**	**.686**	**.649**	**.722**	**.805**	.356
Study 3										
Warmth	**.700**	**.880**	**.893**	**.841**	.370	.419	.130	.225	.429	**-**
Competence	**.815**	.190	.373	.373	**.881**	**.744**	**.802**	**.615**	**.841**	**-**

Re-Analysis Employing Single Items *Warm*, *Competent* and *Moral*, Study 1

In the main text, we present findings for the warmth and competence scales in line with Fiske, Cuddy, Glick and Xu (2002), and treat morality as somewhat distinct from warmth as we believe even though warmth and morality covary in the current studies, they remain conceptually distinct (e.g., Brambilla et al., 2011). Nonetheless, some readers may suspect that our measure of warmth is truly tracking perceptions of morality, corroborated by the above factor analysis. If so, then a re-analysis using only the single items *warm* and *competent* ought to demonstrate diminished effects compared to the scales in the main text. To examine this possibility, we re-conducted the Study 1 analysis using only these items. Results appear quite similar to those in the main text.

We submitted ratings to a 3 (target: self vs. other vs. meta-perceptions) × 2 (decision: harm rejection vs. acceptance) × 2 (personality dimension: warmth vs. competence) repeated measures ANOVA with the first two factors between-subjects and the last factor within-subjects. There was a main effect of target: participants gave higher ratings overall in the self (M = 5.30, SD = 1.45) than the other (M = 4.59, SD = 1.27), or meta-perception conditions (M = 4.71, SD = 1.33), F(2, 194) = 9.13, p < .001, η_p^2 = .09. There was also a main effect of decision: participants rated targets who rejected harm, upholding deontology, higher overall (M = 4.97, SD = 1.03), than targets who accepted harm, upholding utilitarianism (M = 4.21, SD = 1.04), F(2, 194) = 14.14, p < .001, η_p^2 = .17. In addition, there was a main effect of personality dimension: participants gave lower warmth (M = 4.23, SD = 1.73) than competence ratings overall (M = 4.78, SD = 1.33), F(2, 194) = 1.85, p =.160, η_p^2 = .01.

However, these main effects were qualified by a significant two-way interaction between target decision and personality measure, F(1, 194) = 58.48, p < .001, η_p^2 = .23. However, the interactions between target and personality measure, F(2, 194) = .39, p = .691, η_p^2 = .00, and between target and decision, did not reach significance, F(2, 194) = 2.54, p = .081, η_p^2 = .03. More importantly, we again obtained the expected significant three-way interaction, F(2, 194) = 11.92, p < .001, η_p^2 = .11.

Post hoc analyses demonstrated the same pattern of results as those in the main text: Participants in the self-condition rated themselves equally high on warmth when they rejected (*M* = 5.57, *SD* = 1.09) or accepted (*M* = 4.77, *SD* = 1.83) causing harm, *F*(1,194) = 3.01, p =.084, η_p^2 = .02, and equally competent when they rejected (*M* = 5.79, *SD* = 1.05) versus accepted causing harm (*M* = 5.07, *SD* = 1.06), *F*(1,194) = 3.08, *p* = .081, η_p^2 = .02. However, participants in the other-condition replicated the warmth/competence tradeoff found previously: Participants rated Brad higher on warmth when he rejected (*M* = 4.94, *SD* = 1.31), than when he accepted causing outcome-maximizing harm (*M* = 3.63, *SD* = 1.42), *F*(1, 194) = 20.89, *p* < .001, η_p^2 =.10. Conversely, they rated Brad as more competent when he accepted (*M* = 5.07, *SD* = 1.07), than rejected causing outcome-maximizing harm (*M* = 4.19, *SD* = 1.29),-*F*(1, 194) = 12.03, *p* < .001, η_p^2 =.06. Crucially, participants in the meta-perception-condition evinced the same warmth/competence tradeoff as participants in the other-condition: When participants rejected harm, they inferred others would perceive them as warmer (*M* = 5.43, *SD* = 1.95) than when they accepted causing outcome-maximizing harm (*M* = 2.71, *SD* = 1.45), *F*(1, 194) = 22.95, *p* < .001, η_p^2 =.10. In contrast, when they accepted such harm, they inferred that others would perceive them as (slightly) more competent (*M* = 4.83, *SD* = .80) than when they rejected such harm (*M* = 4.36, *SD* = 1.55), although results did not reach conventional levels of significance, *F*(1, 194) = 2.32, *p* = .129, η_p^2 =.01.

Analysis of the Item *Moral*, Studies 1 and 2

As noted in the main text, we assessed the item *moral* in addition to the warmth and competence items. This item evinced patterns quite similar to the warmth composite presented in the main text, and the analysis of the single item *warm* presented above. Nonetheless, we present this item separately, as we believe that warmth/sociability and morality remain conceptually dissociable despite happening to covary here (e.g., Brambilla et al., 2011).

Study 1

Participants in the self-condition rated themselves equally high on morality whether they rejected (*M* = 4.86, *SD* = 1.46) or accepted (*M* = 5.37, *SD* = 1.42) causing harm on moral dilemmas, *F*(1,194) = 1.10, p = .295, η_p^2 = .00. However, participants in the other-condition rated Brad higher on morality when he rejected (*M* = 5.15, *SD* = 1.49), than when he accepted causing outcome-maximizing harm (*M* = 4.24, *SD* = 1.53), *F*(1, 194) = 9.15, *p* < .003, η_p^2 =.05. Crucially, participants in the meta-perception-condition mirrored those in the other-condition: those who rejected harm inferred that others would perceive them as more moral (*M* = 5.64, *SD* = 1.87) than participants who accepted causing outcome-maximizing harm (*M* = 2.71, *SD* = 3.43), *F*(1, 194) = 22.94, *p* < .001, η_p^2 =.10.

Study 2

Participants in the correct-communication condition expected that others would rate them as more moral when they rejected harm, upholding deontology (*M* = 5.06, *SD* = 1.49) than accepted causing harm, upholding utilitarianism (*M* = 3.40, *SD* = 1.68), *F*(1, 182) = 19.70, *p* < .001, η_p^2 = .10. Participants in the error-communication condition demonstrated the opposite pattern: they expected that others would rate them as less moral when they rejected (*M* = 2.94, *SD* = 1.93), rather than accepted, causing harm (*M* = 4.42, *SD* = 1.68), *F*(1, 369) = 22.28, *p* < .001, η_p^2 = .11.

Re-Analysis Employing Kruskal-Wallis & Mann-Whitney tests, Study 1

With uneven cell sizes in Study 1, in some cases the assumption of homogeneity of variance underlying ANOVAs were violated (see below). To account for the impacts of this assumption violation, we conducted non-parametric Kruskal-Wallis & Mann-Whitney Tests, which is more robust to violations of equality of variances (Kruskal & Wallis, 1954; Mann & Whittney, 1947). Results corroborated the analyses in the main text.

Kruskal-Wallis & Mann-Whitney Tests

A Kruskal-Wallis test indicated that the main effect of warmth varied significantly across dilemma decision, H(1) = 28.74, $p < .001$, and target, H(2) = 26.94, $p < .001$. We conducted pairwise comparisons via Mann-Whitney tests to follow up on this finding, while applying a Bonferroni correction by setting alpha to .0167 instead of the traditional .05. This analysis corroborated the analysis in the main text: in the self condition, participants rated themselves equally warm whether rejecting (Mdn = 5.25) or accepting harm (Mdn = 5.25; U = 128.00, p =.430, r = -.13). In the other condition, participants rated Brad higher in warmth when he rejected (Mdn = 5.25) than when he accepted (Mdn = 4.00) causing outcome-maximizing harm (U = 554.00, $p < .001$, r = -.43). Likewise, in the meta-perception condition, participants who rejected harm expected others to rate them as warmer (Mdn = 5.75) than participants who accepted causing outcome-maximizing harm (Mdn = 3.25, U = 104.50, $p < .001$, r = -.27).

A Kruskal-Wallis test indicated that the main effect of competence varied significantly across dilemma decision, H(1) = 10.69, $p < .001$, but not target, H(1) = 3.52, p = .177. We again conducted pairwise comparisons via Mann-Whitney tests with a Bonferroni-correct alpha of .0167 to follow up on this finding. Results indicated that in the self condition, people rated themselves equally competent whether rejecting (Mdn = 5.00) or accepting harm (Mdn = 5.30), U = 147.00, p =.842, r = -.03. In contrast, participants in the other condition rated Brad lower in competence when he rejected (Mdn = 4.60) than accepted (Mdn = 5.20) causing outcome-maximizing harm, U = 708.00, $p < .001$, r = -.31. In the meta-condition, when they accepted such harm, they inferred that others would perceive them as equally competent (Mdn = 4.20) than when they rejected such harm (Mdn = 5.00), U = 171.00, p = .041, r = -.45. Note that these results trending in the hypothesized direction, although they fail to reach Bonferroni-corrected significance levels.

Warmth-Competence Contrasts, Studies 1-3

In the main text, we provide contrasts between warmth across conditions, and separate contrasts between competence across conditions. However, some readers may be interested in comparisons between warmth and competence within condition (see Rom et all., 2017). We present these contrasts here, with the caveat that researchers should exercise caution when interpreting effects across two different scales, as it remains unclear whether participants view these scales as comparable.

Study 1

Participants in the self-condition rated themselves equally high on warmth (*M* = 5.69, *SD* = .33) and competence (*M* = 5.50, *SD* = .28) when they rejected harm, *F*(1,194) = 0.39, *p* = .535, η_p^2 = .001, and equally high on warmth (*M* = 5.12, *SD* = .22) and competence (*M* = 4.85, *SD* = 1.94), when they accepted harm, *F*(1,194) = 1.56, *p* = .213, η_p^2 = .01. However, participants in the other-condition replicated the predicted warmth/competence tradeoff found previously8: When Brad rejected harm (upholding deontology) he received higher warmth (*M* = 5.00, *SD* = 1.19) than competence ratings (*M* = 4.39, *SD* = 1.03), *F*(1,194) = 14.03, *p* < .001, η_p^2 = .07. Conversely, when Brad accepted harm (upholding utilitarianism) he received lower warmth (*M* = 4.03, *SD* = .99) than competence ratings (*M* = 5.12, *SD* = .80), *F*(1, 194) = 39.02, *p* < .001, η_p^2 = .17. Crucially, participants in the meta-perception-condition evinced the same warmth/competence tradeoff as participants in the other-condition: When participants rejected harm they thought others would rate them higher on warmth (*M* = 5.16, *SD* = 1.59) than competence (*M* = 4.38, *SD* = 1.46), *F*(1, 194) = 6.02, *p* = .015, η_p^2 = .03. Conversely, when participants accepted harm they thought others would rate them lower on warmth (*M* = 3.36, *SD* = 1.31) than competence (*M* = 4.89, *SD* = .80), *F*(1, 194) = 69.65, *p* < .001, η_p^2 = .26. For this and other studies with interactions, we also present alternative post-hoc tests in the supplementary material.

Study 2

Post-hoc contrasts largely replicated Study 1 in the correct communication condition: participants who rejected harm (upholding deontology) reported marginally higher warmth (*M* = 5.01, *SD* = 1.47) than competence meta-perception ratings (*M* = 4.52, *SD* = 1.27), *F*(1, 369) = 3.23, *p* = .073, η_p^2 = .01, whereas participants who accepted harm (upholding utilitarianism) reported significantly lower warmth (*M* = 3.48, *SD* = 1.71) than com8petence meta-perception ratings (*M* = 4.83, *SD* = 1.13), *F*(1, 369) = 116.52, $p < .001$, $\eta_p^2 = .24$. As expected, participants in the communication error condition showed the converse pattern: participants who rejected harm (but others thought they accepted harm) reported lower warmth (*M* = 3.04, *SD* = 1.91) than competence meta-perceptions (*M* = 3.66, *SD* = 1.27) , *F*(1, 369) = 7.01, $p = .008$, $\eta_p^2 = .02$, whereas participants who accepted harm (but others thought they rejected harm) reported higher warmth (*M* = 4.38, *SD* = 1.69) than competence meta-perceptions (*M* = 4.00, *SD* = 1.29), *F*(1, 369) = 7.96, $p = .005$, $\eta_p^2 = .02$.

Study 3

Post-hoc tests revealed the same warmth/competence tradeoff as in Studies 1 and 2: participants who rejected harm reported higher warmth (*M* = 4.53, *SD* = 1.61) than competence meta-perception ratings (*M* = 3.92, *SD* = 1.12), *F*(1, 129) = 12.89, $p < .001$, $\eta_p^2 = .09$. Conversely, participants who accepted outcome-maximizing harm reported lower warmth (*M* = 3.22, *SD* = 1.30) than competence meta-perception ratings (*M* = 4.64, *SD* = 1.09), *F*(1, 129) = 68.32, $p < .001$, $\eta_p^2 = .35$.

Chapter 5 – General discussion

In the present dissertation I investigated how internal (e.g., self-perception and meta-perception) and external factors (e.g., a person's moral decision-making and context) influence impression formation and impression management. Chapter 2 identified causal trait theories as a new form of person knowledge, used this construct to explain the emergence of pattern projection, and then offered an explanation for why pattern projection emerges egocentrically. First, people developed more causal trait theories for the self than for someone else, second, the self could more often draw upon behavior from different contexts instead of just one context. That means, by drawing on information of the self across contexts, causal trait theories are not situation-bound, but instead will reflect conclusions from a mini case study based on a variety of contexts. In drawing a connection between behavior in different contexts, people are developing a theory about a person (e.g., the self) that gets exported into how we think about people in general.

In contrast, Chapter 3 focused on how people form impressions of others based on one situation-bound judgment in which personality inferences reflect conclusions only from that one incident. While Chapter 2 focused on conditions in which people project their idiosyncratic self-view in others, Chapter 3 focused on conditions in which different people reach consensus regarding another person's personality independent of their own self-view. I repeatedly found that participants rated targets who rejected causing harm as warmer but less competent than targets who accepted causing harm, because they surmised that harm-rejecting targets engaged in affective processing and they surmised that harm-accepting targets engaged in cognitive processing. This finding was independent of people's own self-assessment, as participants rated themselves high on warmth and competence regardless of their dilemma judgments.

On the surface, it may seem unsurprising to predict that targets who reject causing harm will be perceived as warmer than targets who accept causing harm, as warmth perceptions track benevolence (Fiske et al., 2006) and causing harm is typically not benevolent. Given that people may accept harm in moral dilemmas for benevolent reasons (i.e., to save lives), it is possible that lay people view targets who accept such harm as warmer than people who reject such harm. Similarly, it is difficult to make clear predictions regarding perceptions of competence. As competence tracks effectiveness at realizing one's goals (Fiske et al., 2006), it is possible that lay people perceive targets as competent regardless of whether these targets accept or reject harm, so long the target's choice appears to reflect their (successfully completed) goal. Alternatively, people might view both targets who accept and who reject harm as fairly incompetent, given that both decisions have serious negative consequences. Moreover, it is even possible that people perceive dilemma judgments as uninformative regarding competence if they consider such judgments only informative of the domain of morality. Finally, throughout history there was an association between utilitarianism and emotion, and between deontology and logic (Kagan, 1998). Hence, the present research made clear predictions regarding warmth and competence perceptions that are difficult to arrive at otherwise by putting an emphasis on perceived processing about how a certain decision was reached.

As demonstrated in Chapter 4 participants correctly anticipated how their judgments would influence others' warmth and competence ratings of them—even though, privately, participants rated themselves as high on both warmth and competence regardless of their decision. I then showed that people use meta-perception information to strategically adjust their judgments to create a favorable impression, depending on whether conveying warmth or competence would be more valued in a situation. For example, when participants learned a test focused on emotional competency, participants were more likely to reject causing harm, in contrast, when telling them we were examining logical reasoning abilities, participants were more likely to accept harm. Lastly, I found that decision-makers can use self-presentation strategies to reduce the warmth-competence tradeoff. People who rejected causing harm could offset perceptions of incompetence by describing their decision in terms of logic, whereas decision-makers who accepted outcome-maximizing harm can offset perceptions of coldness by describing their decision in terms of emotions. While much research has focused on intrapsychic processes that drive one to make one or the other decision, I have instead focused on external factors that influence people's moral decision making.

Together, these results show that people infer personality based on their own self-perception and others' harm rejection and acceptance judgments, and that dilemma judgments entail a degree of self-presentation because people have meta-insight into how judgments make them appear. These results suggest that dilemma judgments do not only depend on basic cognitive and affective processes; complex social considerations causally influence dilemma decision-making.

5.2 Limitations and open questions

5.2.1 What is the relevance of external (internal) determinants for impression formation based on causal trait theories (moral dilemma judgments)?

Although, Chapter 2 and 3 looked at internal and external determinants in isolation, I assume that most of the time both factors come into play. The aim of Chapter 2 was to demonstrate the process that leads to egocentric pattern projection. In absence of external information, people's social perception is colored by their own self-perception. However, that does not mean that there is no room for external determinants to factor in. If Jane assumes Joe is artistic but then learns that he was bored last time he went to a gallery she may rate him as less artistic. As attribution research has shown, when forming impressions of others people rely on external cues if they are available. For example, the expressed emotion that comes along with a target's action has shown to influence observers' perception (Krull, Seger, and Silvera, 2008, Ames and Johar, 2009). This also resonates with Study 7, Chapter 4 in which participants were able to reduce the warmth-competence tradeoff when supplementing their decision either with logical or emotional argumentation.

In the same way as there is room for external factors influencing causal trait theories, there is also room for internal factors to influence impression formation following a moral dilemma judgment. Chapter 3 demonstrated a the perception of warmth-competence tradeoff, but the remaining question is how such a process may go about. The current studies were not designed to test the role of the self and thus cannot speak to it. In the case of moral dilemmas, I think it is plausible that people form a lay theory about how certain characteristics in themselves let them reject or accept causing harm. For example, if one experiences strong negative affect in response to contemplating harm while recognizing the logical validity of maximizing outcomes, it is reasonable to infer that others who reject harm experienced strong affect, whereas others who accepted harm engaged in careful deliberation. Consistent with this possibility, recent research suggests that introspective awareness is more common than many psychologists realize (Hahn & Gawronski, 2014), and that one's egocentric perspective easily is brought to mind, whereas adjusting away from it is difficult (Epley, Keysar, Van Boven, & Gilovich, 2004). Thus, in social contexts with limited information, relying on self-insight is a reasonable heuristic for understanding others (Dawes, 1989), so impressions of others are frequently colored by ones' own self-understanding. Hence, lay theories regarding the processes underpinning other's moral dilemma judgments may be grounded in self-perception. This reasoning makes also sense from Chapter 3, Study 4's perspective, as participants were able to predict a dilemma judgment when learning about another person's reasoning.

5.2 Stimulus sampling

One could say that the present research sufferd from biased stimulus sampling. According to a recent paper by Westfall, Judd and Kenny (2012) a small sample of traits is not sufficient to generalize to the population of all traits in general. In Chapter 2 the trait set relied upon eight pre-selected traits, that were all neutral to positive. We conducted multilevel modeling because we expected variation in participants' implicit personality theories for a specific trait based on variation in participants' standing on the relevant two traits, however, it remains to be seen how our results apply to other traits, for example, negative traits. In Chapter 3 we always provided participants with a fixed set of warmth and competence traits taken out from Fiske et al. (2006). Based on our stimulus sampling in the present dissertation following questions remain:

5.2.1 What traits do people spontaneously infer from others dilemma judgments?

Warmth and competence have been found to be the fundamental dimensions of person perception (Fiske et al., 2006), however, there has recently been debated that the single items used for warmth conflate sociability with morality (Goodwin et al., 2014; Leach, Ellemers, & Barreto, 2007). And indeed, Chapter 3 suggests that people believed somewhat different processes might underlie morality ratings as neither perceptions of affective nor cognitive processing mediated the effect of target decision on morality ratings. For example, perceptions of warmth may track perceptions of whether the actor experienced feelings and emotions, whereas perceptions of morality may track whether people acted according to strict moral rules. In a bottom-up approach Goodwin et al. (2014) identified morality and not warmth to be the fundamental dimension of person perception. Our studies, however, were not designed to test the difference between warmth and morality and I suspect that in the context of moral dilemma judgments perceivers draw inferences regarding both warmth and morality from others' moral dilemma judgments. It could be that the current items inflate the correlation between warmth and morality. Future studies could look at which traits participants use spontaneously when inferring impressions from other people's moral dilemma judgments. This would not only shed light on the question which sub dimension of the warmth construct (sociability or morality), but also which sub dimension of the competence construct (intelligence or agency) people represent, when they infer personality inferences based on others moral dilemma decision. Lastly, recent research by Koch, Imhoff, Unkelbach & Alves (2016) suggests that another fundamental dimension of social perception might be progressiveness in beliefs. We did not test for beliefs but future research could test which decision is seen as more progressive and/or whether conservative versus progressive beliefs might be traits participants infer based on another person's moral dilemma judgment.

5.2.2 What traits are spontaneously used in causal trait theories?

The current research only looked at differences between people and not differences between trait pairs. Which are the traits that people are most likely to use to explain one another? The self has privileged access to internal information: feelings, strivings, and personal motivations (e.g., Pronin, 2006). From this perspective, it is plausible to assume that internal and covert traits are most likely to be included in causal trait theories. Another plausible influencing factor could be *trait similarity*. Traits (e.g. bashful and reserved) that are similar to one another are more likely to co-occur in a person, and thus more likely to be used in a causal trait theory. Are positive or negative traits used more frequently? On the one hand, positive traits are more frequent. On the other hand, negative traits capture more attention (Pratto & John, 1991) and are better remembered (Inaba, Nomura, & Ohira, 2005; Ohira, Winton & Oyama, 1998; Ortony, Turner & Antos, 1983). Thus, there might be higher motivation to explain the co-occurrence of negative traits than positive traits. Future research should look into the content of causal trait theories. This might also answer the more interesting question what proper function causal trait theories serve. The value of causal trait theories does not only need to stop with pattern projection but is an interesting endeavor for future research in itself.

5.3 Implications for research on person perception in the moral domain

The current dissertation adds to the growing literature on moral character perception that perceivers not only infer personality based on other's moral decision-making, but also the processes that lead to making one or the other decision. They seem to be aware that affective reactions to harm coincide with the decision to reject harm, and that logical elaboration of outcomes coincides with the decision to accept harm. Various implications have already been addressed in Chapter 2, 3 and 4, I will discuss some broader implications focusing on behavioral consequences and relevance for everyday morality.

The findings of this dissertation suggest far reaching social ramifications, as warmth and competence perceptions influence whom people trust and want to affiliate with. The economist Robert Frank (1988) described how people choose interaction partners in their own lives (at least to some degree) based on judgments of others as being not too coldly rational, and somewhat warmly irrational. For example, choosing a spouse who is competent and instrumental in their dealings may mean that they will be instrumental with us when they find that they can get a better deal elsewhere. This model of commitment shows that another person's attitude or behavior influences our perception of how they are going to treat us. Similarly, warmth and competence perception may influence how we think other people will go about dealing with us.

Theorists disagree regarding the normative status of moral dilemma judgments, with some arguing for the superiority of accepting outcome-maximizing harm in line with utilitarianism (e.g., Greene, 2003) and others championing harm rejection in line with deontology (e.g., Bennis, Medin, & Bartels, 2010). In Chapter 3, participants rated the morality of targets who either accepted or rejected causing harm. Unlike theorists, participants appear to generally agree that targets who reject causing harm are more moral than those who accept causing harm, even when harm maximizes outcomes. As all other target information was held constant, this finding suggests that lay people view characteristically deontological decisions as normatively superior to characteristically utilitarian decisions.

The fact that participants viewed harm rejection as more moral than harm acceptance are in line with the distinction between proscriptive and prescriptive morality drawn by Janoff-Bulman and colleagues (2009). Proscription entails avoiding wrong or harmful behaviors, whereas prescription entails actively helping others. They found that whereas lay people viewed both proscription and prescription as moral, proscriptions were more obligatory than prescription. Therefore, violating a proscription in order to achieve a prescription may be less moral than avoiding a proscription even at the cost of failing a prescription. The current findings also resonate with those of Everett and colleagues (2016) also found that people who make the characteristically deontological judgment are trusted more in economic games and therefore preferred as social interaction partners. In these studies it is usually surmised that people prefer deontologists because they demonstrate strong well-being.

However, Chapter 3 makes an interesting point contrary to research that has found that people generally prefer individuals who make the characteristically deontological decision. My findings rather suggest that preference depends on the situation: people preferred targets who made characteristically utilitarian judgments for the role of hospital manager, a role for which raters prioritize competence over warmth. Thus, depending on what type of social partner people are seeking people may champion the decision to accept harm over the decision to reject harm. In more recent research (Weiss, Rom, & Conway, 2017), we have even found that Brad, when making the characteristically utilitarian judgment is perceived to be more likable, even though less warm and moral. Depending on what kind of social partner people are seeking both decisions can have positive social consequences. This is also corroborated by Chapter 4's finding in which people managed their impressions by selectively choosing the answer that was more adaptive in that situation. If harm-rejection would truly make one appear more positively at all times, participants should not have shifted their responses.

One domain in which one can imagine where it is good to be a utilitarian is leadership. Indeed, competence and agency are seen as leadership qualities (Mumford, 2000). A person who decides to kill one in order to save five may be seen as pragmatic and a good leader, despite lacking empathy and warm character. Hence, previous research that has found that harm-rejecters may be trusted more in economic games has not distinguished different forms of trust. Harm-rejecters will treat us nicely, but may not be able to get the hard things done; harm-accepters may lack empathy, but lead the way when things get tough. Future research may profit from investigating an array of behavioral outcomes by employing economic games. An interesting avenue might be a modified version of a delegation of deception game (Erat, 2013), in which people can delegate displeasing tasks to an agent. Here, people may prefer as cooperation partners agents who make the characteristically utilitarian decision.

5.4 Relevance for everyday life and everyday morality

More generally, Chapter 4's findings can be embedded in broader research on everyday language that has found the way we say something does not only convey a message to the audience, but also conveys information about ourselves and negotiates a relationship with the audience (Pinker, 2007). How we say something influences how we are perceived by others. For example, imagine we want to ask someone to pass on the table salt. When asking for the table salt, we are put in a form of social dilemma: wanting the salt, but not wanting to boss other people around. Thus, we consider the need to convey the message in a way that is polite, and thus add a "please" to our request. Moral dilemmas convey information about the speaker's personality in the same way as language does, and people have meta-insight into what their messages convey. In everyday life, moral dilemmas, albeit not in the severe form of life and death situations, belong to our everyday experiences at all times. Whether we need to decide whether it is morally acceptable to fly, buy clothes, drink alcohol or eat meat, each decision entails a tradeoff. We may think that doing these things makes our life better, but at the same time realize that it is bad for other reasons, such as the health, the environment or animal rights. Our research suggests, for example, that when people decide to eat meat, acknowledging animal suffering may leave a more favorable impression.

5.5 Concluding remarks

The present dissertation looked at internal and external determinants of person perception and impression management, drawing on theory from disparate domains of research. Whereas there is broad interest in moral dilemma decision-making across fields such as neuroscience (e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), experimental philosophy (e.g., Bartels, 2008), and cognitive science (e.g., Moore, Clark, & Kane, 2008), research on person perception primarily stems from the social cognitive tradition (e.g., Weiner, 1985; Fiske, Cuddy, Glick, Xu, 2002). In the current dissertation I integrated theoretical perspectives from these different communities that for too long have hardly cross-talked.

6. References

Abele, A. E., Cuddy, A. J. C., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology, 38*, 1063–1065. doi:10.1002/ejsp.574

Adler, J. M., & McAdams, D. P. (2007). Time, culture, and stories of the self. *Psychological Inquiry, 18*, 97-128.

Ahn., W., Marsh, J. K., Luhmann, C., & Lee, K. (2002). Effect of theory-based feature correlations on typicality judgments. *Memory & Cognition, 30*, 107-118.

Amazon (2011). https://requester.mturk.com/ Retrieved on February 18, 2014.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256.

Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science*, *23,* 861-868. doi:10.1177/0956797611434965

Anderson, C., Ames, D. R., & Gosling, S. D. (2008). Punishing hubris: The perils of overestimating one's status in a group. *Personality and Social Psychology Bulletin, 34,* 90-101. doi:10.1177/0146167207307489

Andersen, S. M., & Ross, L. (1984). Self-knowledge and social inference: I. The impact of cognitive/ affective and behavioral data. *Journal of Personality and Social Psychology*, *46,* 280-293.

Anderson, C. A., & Sedikides, C. (1991). Thinking about people: Contributions of a typological alternative to associationistic and dimensional models of person perception. *Journal of Personality and Social Psychology, 60*, 203-217.

Aquino, A., Haddock, G., Maio, G. R., Wolf, L. J., & Alparone, F. R. (2016). The role of affective and cognitive individual differences in social perception.*Personality and Social Psychology Bulletin*, 0146167216643936.

Aquino, K., & Reed, A. II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83,* 1423-1440. doi:10.1037//0022-3514.83.6.1423

Aristotle (1989/350BC). *Nichomachean ethics*. Oxford: Blackwell.

Aron, A., McLaughlin-Volpe, T., Mashek, D., Lewandowski, G., Wright, S. C., & Aron, E. N. (2004). Including close others in the self. *European Review of Social Psychology, 15*, 101-132.

Arutyunova, K. R., Alexandrov, Y. I., & Hauser, M. D. (2016). Sociocultural Influences on Moral Judgments: East–West, Male–Female, and Young–Old. *Frontiers in Psychology.*

Asch, S. B. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,*258-290.

Asch, S. E., (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 303-314.

Asch, S. E. (1948). The doctrine of suggestion, prestige and imitation in social

psychology. *Psychological Review*, 55, 250–276.

Asch, SE., & Zukier, H. (1984). Thinking about persons. *Journal of Personality and Social Psychology, 46*, 1230-1240.

Baddeley, J., & Singer, J. A. (2010). A loss in the family: Silence, memory, and narrative identity after bereavement. *Memory, 18,* 198-207.

Bakan, D. (1956). *The duality of human existence: Isolation and communion in Western man*. Chicago: Rand McNally.

Barish, K., (Producer), & Pakula, A. J. (Director). (1982). *Sophie's Choice* [Motion picture]. United States: Incorporated Television Company.

Baron, J., & Ritov, I. (2009). Protected values and omission bias as deontological judgments. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.). *Moral judgment and decision making: The psychology of learning and motivation* (Vol. 50, pp. 133–167). SanDiego: Elsevier.

Baron, J., Gürçay, B., Moore, A. B., & Starcke, K. (2012). Use of a Rasch model to predict response times to utilitarian moral dilemmas. *Synthese,* 189, 107–117

Bartels, D. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition, 108*, 381–417. doi:10.1016/j.cognition.2008.03.001

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*, 154-161. doi:10.1016/j.cognition.2011.05.010

Bauer, J. J., & McAdams, D. P. (2010). Eudaimonic growth: Narrative growth goals predict increases in ego development and subjective well-being 3 years later. *Developmental Psychology, 46*, 761-772.

Beauregard, K. S, & Dunning, D. (1998). Turning up the contrast: Self-enhancement motives prompt egocentric contrast effects in social judgments. *Journal of Personality and Social Psychology 74*, 606-621.

Beer, A., & Watson, D. (2008). Asymmetry in judgments of personality: Others are less differentiated than the self. *Journal of Personality, 76*, 535-559. doi:10.1111/j.1467-6494.2008.00495.x

Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science, 5*, 187-202. doi:10.1177/1745691610362354

Bloom, P. (2011). Family, community, trolley problems, and the crisis in moral psychology. *The Yale Review, 99*, 26–43. doi:10.1111/j.1467-9736.2011.00701.x

Borkenau, P., & Liebler, A. (1994). The factor structure of trait ratings depends on the extent of information available to the judges. *European Review of Applied Psychology, 44*, 3-7.

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and sociobiology,13 171-195*. doi:10.1016/0162-3095(92)90032-Y

Boyd, R., & Richerson, P. J. (2005). Solving the puzzle of human cooperation. *Evolution and culture*, 105-132.

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology, 41*, 135-143.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Bruner, J. (2002). *Making stories: Law, literature, life.* New York: Farrar, Straus and Giroux.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3-5. doi:10.1177/1745691610393980

Burnes, B. (2003). *Harry S. Truman: His life and times*. Kansas City, MO: Kansas City Star Books.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and social Psychology, 66*, 840-856. doi:10.1037/0022-3514.66.5.840

Carlson, E. N., & Furr, R. M. (2009). Evidence of differential meta-accuracy: People understand the different impressions they make. *Psychological Science, 20*, 1033-1039. doi: 10.1111/j.1467-9280.2009.02409.x

Carlson, E. N., Vazire, S., & Furr, R. M. (2011). Meta-insight: Do people really know how others see them? *Journal of Personality and Social Psychology, 101*, 831-846. doi: 10.1037/a0024297

Carney, D. R., & Mason, M. F. (2010). Decision making and testosterone: When the ends justify the means. *Journal of Experimental Social Psychology*, *46*(4), 668–671. http://doi.org/10.1016/j.jesp.2010.02.003

Chambers, J. R., Epley, N., Savitsky, K., & Windschitl, P. D. (2008). Knowing too much: Using private knowledge to predict how one is viewed by others. *Psychological Science, 19,* 542-548. doi: 10.1111/j.1467-9280.2008.02121.x

Chapman, L. J., & Chapman, J.P. (1967). Genesis of popular but erroneous diagnostic observations. *Journal of Abnormal Psychology, 72,* 193-204.

Chater, N., & Oaksford, M. (2005). Mental mechanisms: Speculations on human causal learning and reasoning. In K. Fiedler, & P. Juslin (Eds.), Information sampling and adaptive cognition. London: Cambridge University Press.

Chih-Long, Y. (2013). It is our destiny to die: The effects of mortality salience and culture-priming on fatalism and karma belief. *International Journal of Psychology*, *48*, 818-828. doi: 10.1080/00207594.2012.678363

Christensen, J. F., Flexas, A., Calabrese, M., Gut, N. K., & Gomila, A. (2014). Moral judgment reloaded: a moral dilemma validation study. *Frontiers in Psychology, 5*, 1-18. doi:10.3389/fpsyg.2014.00607

Connolly, P. (2009). *Ethics in action: a case-based approach*. John Wiley & Sons.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*, 216-235. doi:10.1037/a0031021

Copeland, B. J. (2014). *Turing: pioneer of the information age*. Oxford University Press.

Critcher, C. R., & Dunning, D. (2009). Egocentric pattern projection: How implicit personality theories recapitulate the geography of the self. *Journal of Personality and Social Psychology, 97*, 1-16.

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, *4*, 308-315.

Critcher, C. R., Dunning, D., & Rom, S. C. (2015). Causal trait theories: A new form of person knowledge that explains egocentric pattern projection. *Journal of personality and social psychology*, *108*, 400.

Crockett, M. J. (2013). Models of morality. Trends in cognitive sciences, 17, 363-366.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92*, 631-648.

doi:10.1037/0022-3514.92.4.631

Cuddy, A. J., Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior*, *31*, 73-98.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*, 353-380.

Cushman, F. (2013). Action, outcome, and value a dual-system framework for morality. *Personality and social psychology review, 17*, 273-292.

Cushman, F., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience, 7*, 269-279.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*, 1082–1089. doi:10.1111/j.1467-9280.2006.01834.x

Dawes, R. M. (1989). Statistical criteria for establishing a truly false consensus effect. Journal of Experimental Social Psychology, 25, 1–17. doi:10.1016/0022-1031(89)90036-X

Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological review*, *94*, 369.

De La Noy, K., Melniker, B. (Producers), & Nolan, C. (Director). (2008). *The Dark Knight* [Motion Picture]. United States: Warner Brothers.

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, *112*, 281-299. doi:10.1017/CBO9780511808098.020

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological bulletin*, *139*, 477. doi:10.1037/a0029065

Dunkel, C. S., & Anthis, K. S. (2001). The role of possible selves in identity formation: A short-term longitudinal study. *Journal of Adolescence, 24,* 765–776.

Dunning, D., & Cohen, G. L. (1992). Egocentric definitions of traits and abilities in social judgment. *Journal of Personality and Social Psychology, 63,* 341–355.

Dunning, D., & Hayes, A. (1996). Evidence for egocentric comparison in social judgment. *Journal of Personality and Social Psychology, 71,* 213–229.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082–1090.

Eagan, J., & Thorne, A. (2010). Life stories of troubled youth: Meanings for a mentor and a scholarly stranger. In K. C. McLean, & M. Pasupathi (Eds.), *Narrative development in adolescence* (pp. 113-129). New York: Springer.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review, 109*(3), 573.

Eagly, A., H., & Karau, S. J. (2002). Role congrueity theory of prejudice towards female leaders. *Psychological Review, 109*, 573-598.

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*. doi:10.1037//0003-066x.54.6.408

Edmonds, D. (2013). *Would you kill the fat man? The trolley problem and what your answer tells us about right and wrong*. Princeton, New Jersey: Princeton University Press.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin, 94*, 100–131. Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87,* 327-339.

Emotionwisegroup (Organization). (2011). Female target conveying happiness [Photograph]. Retrieved April 1st, 2014, from http://www.emotionwisegroup.org/wp/content/uploads/emotipedia/emotipedia/

Emotionwisegroup (Organization). (2011). Male target conveying happiness [Photograph]. Retrieved April 1st, 2014, from http://www.emotionwisegroup.org/wp-content/uploads/emotipedia/emotipedia/

Epley, N., & Dunning, D. (2000). Feeling "Holier than thou": Are self-serving assessments produced by errors in self or social prediction? *Journal of Personality and Social Psychology*, *79,* 861-875. doi: 10.1037/0022-3514.79.6.861

doi: 10.1037/0022-3514.87.3.327

Epley, N., Morewedge, C., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social psychology, 40,* 760-768.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87,* 327–339. doi: 10.1037/0022- 3514.87.3.327

Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General, 145*, 772.

Faul, F., Erdfelder, E., Lang, A.-G., &; Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. doi:10.3758/BF03193146

Feinberg, M., Willer, R., Stellar, J., & Keltner, D. (2012). The virtues of gossip: reputational information sharing as prosocial behavior. *Journal of Personality and Social Psychology*, *102*, 1015. doi:10.1037/a0026650

Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological science*, *25,* 656-664. doi:10.1177/0956797613510184

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

Fischer, P., Greitemeyer, T., Pollozek, F., & Frey, D. (2006). The unresponsive bystander: Are bystanders more responsive in dangerous emergencies? *European Journal of Social Psychology*, *36*, 267-278. doi:10.1002/ejsp.297

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2006). Universal dimensions of social cognition: Warmth and Competence. *Trends in Cognitive Sciences, 11*, 77-83. doi:10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878-902.

doi: 10.1037/0022-3514.82.6.878

Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis) respecting versus (dis) liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, *55*, 473-489.

Fivush, R., Bohanek, J. G., & Marin, K. (2010). Patterns of family narrative co-construction in relation to adolescent identity and well-being. In K. C. McLean, & M. Pasupathi (Eds.), *Narrative development in adolescence* (pp. 45-63). Springer.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5,* 5-15. doi:10.1093/0199252866.003.0002

Friedman, L., Fleishman, E. A., & Fletcher, J. M. (1992). Cognitive and interpersonal abilities related to the primary activities of R&D managers. *Journal of Engineering and Technology Management*, *9*, 211-242.

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in moral judgments: A process dissociation meta-analytic reanalysis. *Personality and Social Psychology Bulletin*.

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender Differences in Responses to Moral Dilemmas A Process Dissociation Analysis. *Personality and Social Psychology Bulletin*, 0146167215575731.

Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Sposta, S., Zago, S., ... Priori, A. (2010). Gender-related differences in moral judgments. *Cognitive Processing*, *11*, 219–226. doi:10.1007/s10339-009-0335-2

Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R., & Hütter, M. (2015). Moral dilemma judgments: Disentangling deontological inclinations, utilitarian inclinations, and general action tendencies. In J. P. Forgas, P. A. M. Van Lange, & L. Jussim (Eds.), *Social psychology of morality*. New York: Psychology Press.

Giangreco, D. M. (1997). Casualty projections for the U.S. invasion of Japan, 1945-1946: Planning and policy implications. *The Journal of Military History, 61*, 521-82.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology, 75,* 332-346. doi:10.1037/0022-3514.75.2.332

Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. PLOS ONE, 8, 1-9. doi:10.1371/journal.pone.0060418

Gold, N., Colman, A. M., & Pulford, B. D. (2014). Cultural differences in responses to real-life and hypothetical trolley problems. Judgment and Decision Making, 9, 65-76.

Gold, N., Pulford, B. D., & Colman, A. M. (2015). Do as I say, don't do as I do: Differences in moral judgments do not translate into differences in decisions in real-life trolley problems. Journal of economic psychology, 47, 50-61.

Goldings. (1954). On the avowal and projection of happiness. Journal of Personality, 23, 30–47.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. Journal of Personality and Social Psychology, 106, 148-168. doi:10.1037/a0034726

Greene, J. D. (2003). From neural 'is' to moral 'ought': What are the moral implications of neuroscientific moral psychology? Nature Reviews, Neuroscience, 4, 847-850. doi:10.1038/nrn1224

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*, 364-371. doi:10.1016/j.cognition.2009.02.001

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389-400. doi:10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105-2108. doi:10.1126/science.1062872

Grice, H. P. (1989). Studies in the Way of Words. Cambridge, MA: Harvard University Press.

Habermas, T., & Buck, S. (2000). Getting a life: The emergence of the life story in adolescence. *Psychological Bulletin, 126*, 748-769.

Hahn, A., & Gawronski, B. (2014). Do implicit evaluations reflect unconscious attitudes? *Behavioral and Brain Sciences*, *37*, 28–29. doi:10.1017/S0140525X13000721

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834. doi:10.1037//0033-295X.108.4.814

Hamilton, D.L., Katz, L.B., & Leirer, V.O. (1980). Cognitive representation of personality impressions: Organizational processes in first impression formation. *Journal of Personality and Social Psychology, 39*, 1050-1063.

Hampson, S. E. (1998). When is an inconsistency not an inconsistency? Trait reconciliation in personality description and impression formation. *Journal of Personality and Social Psychology, 74*, 102–117.

Han, H., Glover, G. H., & Jeong, C., (2014). Cultural influences on the neural correlate of moral decision making processes. *Behavioral Brain Research, 259*, 215-228. doi:10.1016/j.bbr.2013.11.012

Hess, N. H., & Hagen, E. H. (2006). Psychological adaptations for assessing gossip veracity. *Human Nature, 17*, 337-354. doi:10.1007/s12110-006-1013-z

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science, 345*, 1340-1343. doi:10.1126/science.1251560

Holmes, D. S. (1981). Existence of classical projection and the stress reducing function of attributive projection: A reply to Sherwood. *Psychological Bulletin, 90,* 460–466.

Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology, 4,* 386. doi:10.3389/fpsyg.2013.00386

Inbar, Y., Pizarro, D., & Cushman, F. (2012). Benefitting from misfortune: When harmless actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin, 38*, 52-62. doi:10.1177/0146167211430232

Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology, 96*, 521-537. doi:10.1037/a0013779

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, 96*, 521-537. doi:10.1037/0022-3514.89.6.899

Judd, C. M., Kenny, D. A., & Krosnick, J. A. (1983). Judging the positions of political candidates: Models of assimilation and contrast. *Journal of Personality and Social Psychology, 44,* 952–963.

Kagan, (1998). *Normative Ethics*. Westview Press: Boulder, CO.

Kahane, G. (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social neuroscience, 10*(5), 551-560.

Kahane, G., Everett, J. A. C., Earp, B. D., Farias, M., &amp; Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition, 134*, 193-209. doi:10.1016/j.cognition.2014.10.005

Kant, I. (1785/1959). *Foundation of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis: Bobbs-Merrill.

Kaplan, S. A., Santuzzi, A. M., & Ruscher, J. B. (2009). Elaborative metaperceptions in outcome-dependent situations: The diluted relationship between default self-perceptions and metaperceptions. *Social Cognition, 27*, 601-614. doi: 10.1521/soco.2009.27.4.601

Katz, D., & Allport, F. (1931). *Students' attitudes*. Syracuse, NY: Craftsman Press.

Kawai, N., Kubo, K., & Kubo-Kawai, N. (2014). "Granny dumping": Acceptability of sacrificing the elderly in a simulated moral dilemma. *Japanese Psychological Research, 56*, 254-262.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57*, 227-254.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation*. Lincoln: University of Nebraska Press.

Kenny, D. A., & DePaulo, B. M. (1993). Do people know how others view them? An empirical and theoretical account. *Psychological Bulletin, 114,* 145-161.

doi:10.1037/0033-2909.114.1.145

Kim, M.P., Rosenberg, S. (1980). Comparison of two structural models of implicit personality theory. *Journal of Personality and Social Psychology, 38*, 375–89.

King, L. A., Burton, C. M., & Geise, A. C. (2009). The good (gay) life: The search for signs of maturity in the narratives of gay adults. In P. L. Hammack, & B. J. Cohler (Eds.), *The story of sexual identity: Narrative perspectives on the gay and lesbian life course* (pp. 375-396). Oxford University Press.

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. Journal of personality and social psychology, 110(5), 675.

Kohlberg, L. (1969). Stage and sequence: The cognitive–developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. (pp. 347–480). Chicago, IL: Rand McNally.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature 446*, 908-911. doi:10.1038/nature05631

Krebs, D. L. (2011). The evolution of a sense of morality. *Creating consilience*, 299-317.

Kreps, T. & Monin, B. (2014). Core values vs. common sense: Consequentialist views appear less rooted in morality. *Personality and Social Psychology Bulletin, 40*, 1529-1542. doi:10.1177/0146167214551154

Kruger, J., & Gilovich, T. (2004). Actions, intentions, and self-assessment: The road to self-enhancement is paved with good intentions. *Personality and Social Psychology Bulletin*, *30*(3), 328-339.

Krueger, J., & Stanke, D. (2001). The role of self-reference and other referent knowledge in perceptions of group characteristics. *Personality and Social Psychology Bulletin, 27,* 876–888.

Kunda, Z., Miller, D. T., & Claire, T. (1990). Combining social concepts: The role of causal reasoning. *Cognitive Science, 14,* 551-577.

Kundu, P., & Cummins, D. D. (2013). Morality and conformity: The Asch paradigm applied to moral decisions. *Social Influence*, *8,* 268-279. doi:10.1080./15534510.2012.727767

Laing, R. D., Phillipson, H., & Lee, A. R. (1966). *Interpersonal perception: A theory and a method of research*. Oxford: Springer.

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

Personality and Social Psychology, 93, 234-249. doi:10.1037/0022-3514.93.2.234

Leary, M. R. (1989). Self-presentational processes in leadership emergence and effectiveness.

Leary, M. R. (1995). *Self-presentation: Impression management and interpersonal behavior*. Brown & Benchmark Publishers.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological bulletin*, *107*(1), 34.

Liu, B. S., & Ditto, P. H. (2013). What dilemma? Moral evaluation shapes factual belief. *Social Psychological and Personality Science, 4*, 316-323. doi:10.1177/1948550612456045

López, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development, 63*, 1070–1090.

Lucas, B. J., & Galinsky, A. D. (2015). Is utilitarianism risky? How the same antecedents and mechanism produce both utilitarian and risky choices. *Perspectives on Psychological Science*, *10*(4), 541-548.

Lucas, B. L., & Livingstone, R. W. (2014). Feeling socially connected increases utilitarian choices in moral dilemmas. *Journal of Experimental Social Psychology, 53*, 1–4.

doi:10.1016/j.jesp.2014.01.011

Malle, B. F., Knobe, J., & Nelson, S. (2007). Actor-observer asymmetries in behavior explanations: New answers to an old question. *Journal of Personality and Social Psychology, 93*, 491–514.

Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224-253.

McAdams, D. P. (1985). *Power, intimacy, and the life story: Personological inquiries into identity*. New York: Guilford Press. McAdams, D. P. (2001). The psychology of life stories. *Review of General Psychology, 5*, 100–122.

McAdams, D. P. (1995). What do we know when we now a person? *Journal of Personality, 63*, 363-396.

McAdams, D. P. (2001). The psychology of life stories. *Review of General Psychology, 5*, 100-122.

McAdams, D. P., & McLean, K. C. (2013). Narrative identity. *Current Directions in Psychological Science, 22*, 233-238.

McNorgan, C. M., Kotack, R. A., Meehan, D. C., & McRae, K. (2007). Feature-feature causal relations and statistical co-occurrences in object concepts. *Memory & Cognition, 35*, 418-431.

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, *121,* 277-301.

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *TRENDS in Cognitive Sciences, 11*, 143-152. doi:10.1016/j.tics.2006.12.007

Milgram, S. (1963). Behavioral Study of obedience. *The Journal of abnormal and social psychology*, *67*(4), 371.

Mill, J. S. (1861/1998). *Utilitarianism*. In R. Crisp (Ed.), New York: Oxford University Press.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science, 19*, 549-57. doi:10.1111/j.1467-9280.2008.02122.x

Mumford, M. D., Zaccaro, S. J., Connelly, M. S., & Marks, M. A. (2000). Leadership skills: Conclusions and future directions. *The Leadership Quarterly*, *11*, 155-170.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in cognitive science: Vol. 2. Theory and applications* (pp. 23-45). Chichester, U.K.: Ellis Horwood.

Navarrete, C. D., McDonald, M. M., Mott, M. L. & Asher, B. (2011). Virtual morality: emotion and action in a simulated three-dimensional "trolley problem". *Emotion, 12*, 364–370. doi:10.1037/a0025561

Nichols, S., Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition, 100,* 530-542. doi:10.1016/j.cognition.2005.07.005

Pals, J. L. (2006). Constructing the “springboard effect”: Causal connections, self-making, and growth within the life story. In D. P. McAdams, R. Josselson, and Lieblch (Eds.), *Identity and story: Creating self in narrative* (pp. 175–199). Washington, DC: American Psychological Association Press.

Park, B. (1986). A method for studying the development of impressions of real people. *Journal of Personality and Social Psychology, 51,*907-917.

Park, B., DeKray, M. L., & Kraus, S. (1994). Aggregating social behavior into person models: Perceiver-induced consistency. *Journal of Personality and Social Psychology*, *66*, 437-459.

Pasupathi, M., & Wainryb, C. (2010). On telling the whole story: Facts and interpretations in autobiographical memory narratives from childhood through midadolescence. *Developmental Psychology, 46*, 735-746.

Peeters, G. (1983). Relational and informational pattern in social cognition. In W. Doise & S. Moscovici (Eds.), *Current Issues in European Social Psychology* (201–237). Cambridge: Cambridge University Press.

Pinker, S. (2008, January 13). The moral instinct. *The New York Times*. Retrieved from http://www.nytimes.com

Pizarro, D. A. (2000). Nothing more than feelings? The role of emotions in moral judgment. *Journal for the Theory of Social Behavior, 30*, 355-375.

Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The Social Psychology of Morality: Exploring the Causes of Good and Evil* (pp. 91–108). Washington, DC: American Psychological Association.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717-731.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879-891. doi:10.3758/BRM.40.3.879

Prentice, D. A. (1990). Familiarity and differences in self- and other-representations. *Journal of Personality and Social Psychology, 59*, 369-383.

Prinz, J. (2006). The emotional basis of moral judgment. *Philosophical Explorations, 9*, 29-43.

Pronin, E. (2008). How we see ourselves and how we see others. *Science, 320*, 1177-1180.

doi: 10.1126/science.1154199

Pronin, E., Kruger, J., Savitsky, K., & Ross, L. (2001). You don't know me, but I know you: The illusion of asymmetric insight. *Journal of Personality and Social Psychology, 81*, 639-656.

Reese, E., Yan, C., Jack, F. & Hayne, H. (2010). Emerging identities: Narrative and self from early childhood to early adolescence. In K. C. McLean, & M. Pasupathi (Eds.), *Narrative development in adolescence* (pp. 23-43). New York: Springer.

Reis, H. T., & Gruzen, J. (1976). On mediating equity, equality, and self-interest: The role of self-presentation in social exchange. *Journal of Experimental Social Psychology*, *12*(5), 487-503.

Risen, J. L., & Gilovich, T., Dunning, D. (2007). One-shot illusory correlations and stereotype formation. *Personality and Social Psychology Bulletin, 33, 1492- 1502.*

Rockler, M. (2007). Presidential decision-making. *Philosophy Now*, *64*, 18-19.

Rom, S. C., Weiss, A., & Conway, P. (2017). Judging those who judge: Perceivers infer the roles of affect and cognition underpinning others' moral dilemma responses. *Journal of Experimental Social Psychology*, *69*, 44-58.

Rosenberg, S, (1976). New approaches to the analysis of personal constructs in person perception. In A. W Lanfield (Ed.), *Nebraska symposium on motivation* (pp. 179-242). Lincoln: University of Nebraska Press.

Rosenberg, S., Nelson, C. & Vivekananthan, P. S., (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*, 283-294. doi:10.1037/h0026086

Ross, L., Greene, D., & House, P. (1977). The false consensus phenomenon: An attributional bias in self-perception and social perception processes. *Journal of Experimental Social Psychology, 13*, 279-301.

Royzerman, E.B., Landy, J. F., & Leeman, R. F. (2014). Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. *Cognitive Science, 39*, 1-28. 10.1111/cogs.12136

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of personality and social psychology*, *77*, 1004.

doi:10.1037/0022-3514.77.5.1004

Sarbin, T. R., & Allen, V. L. (1968). Role theory. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (pp. 488–567). Reading, MA: Addison-Wesley.

Schnall, S., Roper, J., & Fessler, D. M. T. (2010). Elevation leads to altruism, above and beyond general positive affect. *Psychological Science, 21*, 315-320. doi:10.1177/0956797609359882

Schneid, E. D., Carlston, D. E., & Skowronski, J. J. (2015). Spontaneous evaluative inferences and their relationship to spontaneous trait inferences. *Journal of Personality and social Psychology, 108*, 681-696. doi:10.1037/a0039118

Sedikides, C. (1993)., *Assessment, enhancement, and verification determinants of the self-evaluation process. Journal of Personality and Social Psychology, 65,* 317–338.

Sedikides, C., & Anderson, C. A. (1994). Causal perceptions of intertrait relations: The glue that holds person types together. *Personality and Social Psychology Bulletin*, *20*, 294-302.

Sedikides, C., & Strube, M.J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. *Advances in Experimental Social Psychology, 29*, 209-269.

Sherif, M. A. (1935). A study of some social factors in perception, *Archives of Psychology*, 27,

1– 60.Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making*, *4*(6), 479.

Singer, P. (2004). *The president of good and evil*. London: Granta.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file drawer. *Journal of Experimental Psychology: General, 143*, 534-547. doi:10.1037/a0033242

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology, 88,* 895. doi:10.1037/0022-3514.88.6.895

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*, 164-182.

Sterba, J. P. (1988). *How to make people just*. Totowa, NJ: Rowman and Littlefield.

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

Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition, 119,* 454-458. doi:10.1016/j.cognition.2011.01.018

Szekely, R. D., Opre, A., & Miu, A. C. (2015). Religiosity enhances emotion and deontological choice in moral dilemmas. *Personality and Individual Differences, 79*, 104-109. doi:10.1016/j.paid.2015.01.036

Szekely, R. D. & Miu, A. C. (2014). Incidental emotions in moral dilemmas: The influence of emotion regulation. *Cognition & Emotion, 29*, 1-12. doi:10.1080/02699931.2014.895300

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology, 4*, 1-8. doi:10.3389/fpsyg.2013.00250

Tetlock, P. E. (2000). The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78,* 853-870. doi:10.1037//0022-3514.78.5.853

Trémolière, B., & Bonnefon, J. F. (2014). Efficient kill–save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin, 40*, 923-930.

Trémolière, B., Kaminski, G., & Bonnefon, J. F. (2015). Intrasexual competition shapes men’s anti-utilitarian moral decisions. *Evolutionary Psychological Science*, *1*, 18-22.

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. Annu. Rev. Psychol., 59, 329-360.

Uhlmann, E. L., Pizarro, D., A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making, 4*, 476-491.

Uhlmann, E. L., Zu, L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition, 126,* 326-334. doi:10.1016/j.cognition.2012.10.005

Valdesolo, P. & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17*, 476–477. doi:10.1111/j.1467-9280.2006.01731.x

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*, 281-300.

doi: 10.1037/a0017908

Von Baeyer, C. L., Sherk, D. L., & Zanna, M. P. (1981). Impression management in the job interview: When the female applicant meets the male (chauvinist) interviewer. *Personality and Social Psychology Bulletin*, *7*(1), 45-51.

Vorauer, J. D., & Claude, S. (1998). Perceived versus actual transparency of goals in negotiation. *Personality and Social Psychology Bulletin, 24*, 371-385.

doi:10.1177/0146167298244004

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*, 548.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. Journal of Experimental Psychology: General, 143, 2020.

Whitlock, C. (2006, February 16). German court overturns law allowing hijacked airliners to be shot down. *Washington Post Foreign Service.* Retrieved from http://www.washingtonpost.com

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, 37*, 395-412.

doi:10.1037/0022-3514.37.3.395

Winkielman, P., & Schwarz, N. (2001). How pleasant was your childhood? Beliefs about memory shape inferences from experienced difficulty of recall. *Psychological Science*, *12*(2), 176-179.

Winterbotham, F. W. (1974). The Ultra Secret (New York, 1974). *Ronald Lewin, Ultra Goes to War (New York, 1978).*

Wojciszke**,** B. (1994). Multiple meanings of behavior: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology, 67*, 222-232. doi:10.1037/0022-3514.67.2.222

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. *Advances in experimental social psychology*, *46*, 55-123.

Wright, S. C., Aron, A., & Tropp, L. R. (2002). Including others (and groups) in the self: Self-expansion and intergroup relations. In J. P. Forgas & K. D. Williams (Eds.), *The social self: Cognitive, interpersonal and intergroup perspectives* (pp. 343-363). Philadelphia: Psychology Press.