Shared Reality or Shared Illusions? Evaluating Moral Impressions

by

Maxwell Barranti

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Psychology
University of Toronto

© Copyright by Maxwell Barranti 2019
Abstract

Moral impressions are some of the most consequential opinions people have about themselves and others. Morality is at the core of identity (Strohminger & Nichols, 2014), and drives many interpersonal interactions such as cooperation (Delgado, Frank, & Phelps, 2005) and affiliation (Bukowski & Sippola, 1998). Yet, little is known about the extent to which these moral impressions are grounded in reality or if they are only held as idiosyncratic impressions. The current investigation evaluated how people see their own and other’s moral character, if these moral impressions are shared, and if a shared understanding of moral character is adaptive. First, I developed and tested a measurement model for assessing self- and other-impressions of morality which removes a global evaluative bias from moral impressions (Studies 1 and 2). Second, I evaluated if self- and other-impressions are grounded in a shared social reality (i.e. self-other agreement and inter-judge consensus) and/or grounded in observable behavior (Study 3). Third, I evaluated if sharing a social reality for moral impressions has interpersonal consequences for the self (Study 4). This work sheds light on the extent to which morality is in the eye of the beholder and the adaptiveness of holding shared moral impressions. Additionally, this work has implications for the assessment of moral character and the adaptiveness of self-knowledge.
Acknowledgments

I would like to thank the many mentors that have supported me over the years. Without their help none of this would have been possible. To Erika, Mike, Uli, Matt, Stéphane, Jenny, Yoel, Tera, and Kate – thank you!
# Table of Contents

Acknowledgments .................................................................................................................. iii

Table of Contents .................................................................................................................. iv

List of Tables ......................................................................................................................... viii

List of Figures ......................................................................................................................... ix

List of Appendices ................................................................................................................ x

Chapter 1 General Introduction ........................................................................................... 1

1 Importance of moral impressions ....................................................................................... 1

1.1 Self-impressions ............................................................................................................... 1

1.2 Other-impressions .......................................................................................................... 2

2 How do accurate moral impressions form? ........................................................................ 3

3 Challenges to forming accurate moral impressions .......................................................... 5

3.1 What is morally relevant? ............................................................................................... 5

3.2 Moral traits are not “good” traits .................................................................................. 8

3.2.1 Evaluative .................................................................................................................. 8

3.2.2 Internal ..................................................................................................................... 9

3.3 Asymmetries in achieving accuracy .............................................................................. 10

4 Ways of assessing accuracy .............................................................................................. 11

4.1 Self-other agreement and consensus ............................................................................ 11

4.2 Behavioral Criteria ......................................................................................................... 12

Chapter 2 Measuring Moral Impressions ............................................................................ 14

1 Study 1 methods ................................................................................................................ 17

1.1 Participants ..................................................................................................................... 17

1.2 Procedure ....................................................................................................................... 18

1.3 Materials ......................................................................................................................... 18
1.3.3 Public goods game ................................................................. 41
1.4 Daily behaviors ........................................................................... 42
2 Study 3 results ................................................................................ 43
  2.1 Inter-judge consensus................................................................. 43
  2.2 Self-other agreement ................................................................. 44
  2.3 Links between moral impressions and behavior ......................... 45
3 Study 3 Discussion ........................................................................ 48
Chapter 4 What are the Consequences of Sharing Moral Impressions? .... 50
  1 Is a shared reality adaptive? ......................................................... 50
  2 Using RSA to test associations with shared reality ....................... 51
3 Study 4 methods ........................................................................... 53
  3.1 Participants ............................................................................... 53
  3.2 Materials and procedure ........................................................... 53
  3.3 Analytic approach .................................................................... 54
    3.3.1 Controlling for evaluative bias ........................................... 56
4 Study 4 results ............................................................................. 57
5 Study 4 discussion ........................................................................ 62
Chapter 5 General Discussion .......................................................... 65
  1 Extent of evaluative bias ............................................................... 65
  2 Inter-judge consensus ................................................................. 65
  3 Self-other agreement ................................................................. 66
  4 In lab and daily behaviors ............................................................ 66
  5 Is sharing reality adaptive? .......................................................... 67
  6 Implications for the study of moral psychology ........................... 67
  7 Limitations and future directions ................................................. 68
  8 Conclusion .................................................................................. 69
List of Tables

Table 1. Moral Domains According to Various Scholars ................................................................. 6

Table 2. List of Items Used To Measure Moral Impressions and Evaluative Bias ....................... 19

Table 3. Quantifying the Relationship Between Evaluative Bias and Moral Impressions at the Factor and Item Levels ........................................................................................................... 21

Table 4. Amount of Variance in Moral Impressions Attributable Morality, Bias, Domain, and Error ........................................................................................................................................ 31

Table 5. Correlation Matrix Among Socially Desirable Responding Measures, Moral Factor, and Evaluative Bias .................................................................................................................................. 33

Table 6. Model Fit and ΔModel Fit for Pattern Invariance Models .............................................. 36

Table 7. Polynomial Slopes and Response Surface Coefficients for Moral Impressions, Extraversion, and Agreeableness [95% CI] ........................................................................................................ 59
List of Figures

Figure 1. Models Used to Partition Evaluative Bias from Moral Impressions .......................... 20

Figure 2. Models Used to Test for Invariance ........................................................................ 24

Figure 3. Illustration of Nested Models for Self-Other Pattern Invariance .......................... 34

Figure 4. Model of inter-judge consensus ................................................................................ 43

Figure 5. Structural Equation Model of Self-other Agreement ............................................... 46

Figure 6. Relationships Among Behaviors, Self-, and Other-reports of Morality ....................... 47

Figure 7. Visualization of Response Surfaces That Would Support Hypotheses ..................... 55

Figure 8. Response Surfaces of Moral Impressions, Extraversion, and Agreeableness ............. 60
List of Appendices

Appendices A: List of Scales Included in Study 1 ................................................................. 81
Appendices B: List of Scales Included in Study 2 .................................................................. 82
Appendices C: Items in Socially Desirable Responding Scales ............................................. 83
Appendices D: List of Scales and Tasks in Study 3 ............................................................... 84
Chapter 1
General Introduction

Moral impressions are perhaps among the most consequential and fundamental impressions people hold about themselves and others. Indeed, there are many reasons to believe moral impressions are important and at the heart of person perception (Brambilla, & Leach, 2014; Goodwin, 2015). Moral character lays at the core of identity (Strohminger & Nichols, 2014), people preferentially seek out information about other’s morality (Brambilla, Rusconi, Sacchi, & Cherubini, 2011), people use moral impressions to make important decisions (e.g., hiring decisions; Sackett & Harris, 1984), and positive moral impressions confer social benefits (e.g., liking and status; Ellemers, Pagliaro, Barreto, & Leach, 2008).

As important as moral impressions are in our everyday lives, people face many challenges to forming accurate moral impressions about themselves and others. This is in part because forming accurate impressions is difficult in general. But additionally, moral impressions are particularly evaluative (i.e. it is obviously desirable to be seen as moral) and at least somewhat internal. These features exacerbate the challenge of being accurate about morality. Given that moral impressions are important to people’s lives, but at the same time difficult to judge, moral impressions are a particularly interesting area to assess self-knowledge and accuracy. The overarching goal of this work is to evaluate the extent to which people’s moral impressions are grounded in reality and the consequences of failing to hold grounded moral impressions. This broad goal will be tackled in several parts. First, I will develop a measurement model to assess moral impressions and the extent to which they are clouded by evaluative bias. Second, I will evaluate if moral impressions are shared among perceivers and associated with observable behavior. Third, I’ll evaluate if sharing moral impressions with other perceivers is adaptive.

1 Importance of moral impressions

1.1 Self-impressions

The beliefs we hold about our own moral character – moral self-impressions – are among the most important beliefs we hold about ourselves (Wojciszke, 2005). Indeed, people’s beliefs about their own morality lie at the heart of their identity (Heiphetz, Strohminger, & Young, 2017). For example, changing aspects of our moral selves is viewed as fundamentally changing a
person, more so than non-moral aspects (Strohminger & Nichols, 2014). When people are asked about what traits comprise the ideal person, moral traits are the defining features of an ideal person (Cottrell, Neuberg, & Li, 2007). Even after death, loved ones emphasize the morality of the recently deceased (Goodwin, Piazza, & Rozin, 2014). People value a moral self-image (Jordan & Mullen, 2011; Monin & Jordan, 2009) and have a strong desire to see themselves as moral (Mazar, Amir, & Ariely, 2008). Given how much people care about their own morality it stands to reason that holding moral self-perceptions may be a fundamental psychological need (Prentice, Jayawickreme, Hawkins, Hartley, Furr, & Fleeson, 2018). Taken together, this suggests that moral self-perceptions are an important part of people’s lives.

1.2 Other-impressions

It isn’t just our own morality that people care deeply about. People also care deeply about the morality of others (Pizarro, & Tannenbaum, 2012). This is evidenced by the amount of time and effort people put into seeking and sharing information about the moral character of others. People preferentially seek out morally relevant information more so than non-moral information when forming an impression of others (Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Wojciszke, Bazinska, & Jaworski, 1998). Further, and they regularly share information about the moral character of people they know. For example, people regularly gossip about morality (Feinberg, Willer, Stellar, & Keltner, 2012; Peters & Kashima 2015). A conservative estimate is that on any given day there is about a 15% chance to learn about moral acts of others (Hofmann, Wisneski, Brandt, & Skitka, 2014). In essence, people exert effort into learning the morality of others.

Part of why people try to understand the other’s morality is because they believe it is useful information for guiding important interpersonal decisions. Perhaps unsurprisingly, people use their impressions about other’s moral character to inform a wide variety of decisions. As such, moral impressions carry consequence. Moral character impressions affect who people like and trust (Goodwin et al., 2014). Friendships and relationships are sought and ended based on moral judgments (Van’t Wout & Sanfey, 2008). Decisions about whom to trust with valuable resources (e.g. money) are informed by moral character judgments (e.g. money; De Bruin & Van Lange, 1999; Delgado, Frank, & Phelps, 2005). People avoid working with people who are uncooperative. For example, people prefer to work with other’s when they have seen them make
large contributions to a public good (Rockenbach & Milinski, 2011). Leaders are elected based on constituent impressions of their morality. For example, voters prefer candidates that appear trustworthy (Chen, Jing, & Lee, 2014). All of this evidence suggests that being seen as a moral person has positive interpersonal consequences.

In essence, people care deeply about their own morality and exert lots of effort to understand the moral character of others. The well documented dominance of moral impressions in our social lives has driven scholars to regard morality as a fundamental dimension of person perception (Brambilla & Leach, 2014, Goodwin et al., 2014; Wojciszke, 2005). And yet, despite the importance of moral impressions in our lives and its fundamental role in person perception, current work has not adequately addressed several fundamental questions about moral character impressions. Are our moral impressions grounded in reality? Is holding realistic moral impressions adaptive? To address these fundamental questions, it is important to understand the process by which accurate moral impressions form and the challenges that stand in the way of forming accurate moral impressions.

2 How do accurate moral impressions form?

The process by which people form accurate moral impression is no different than the process for any personality trait. The Realistic Accuracy Model (RAM) outlines the successive steps that must happen for a person to form an accurate impression (Funder 1995, 2012). The first two steps are largely about the person being judged. For there to be any hope of achieving accurate personality judgments of morality there must be relevant cues about a person’s morality, which are then made available to the judge. The second two steps are largely about the judge. The judge must successfully detect available cues and then utilize those cues in the appropriate way. By successfully navigating this process, people may become accurate judges of any trait – including morality.

When it comes to morality, there must be morally relevant cues in order for people to form accurate moral judgements. In other words, a person can’t simply be inert. There are many cues that are likely relevant to a person’s morality. Overt pro-social or anti-social behaviors can serve as cues of morality. Pro-social behaviors such as volunteering time, donating blood, giving money to the poor are widely regarded as moral (Batson, Thomsom, & Chen, 2002; Staub, 1978). Other relevant cues may come into existence when people respond to the immoral
behaviors of others. For example, punishing wrongdoers is a relevant cue to trustworthiness (Jordan, Bloom, & Rand, 2016). Or relevant cues might come into existence when people react to their own immoral behaviors. For example, people show signs of guilt when they regret having acted immorally (Cohen, Panter, & Turan, 2012; Tangney, Wagner, & Gramzow, 1992; Wicker, Payne, & Morgan, 1983). As such, people’s behaviors and their responses to behavior can serve as relevant cues to a person’s morality.

These moral cues will not help people form accurate impressions unless the cues are made available to judges. For example, someone who regularly but anonymously donates to charity would deprive those judges around them of this cue to their morality. It is perhaps not too surprising then that people often make cues about their morality available as a way of signaling to others that they are in fact moral. For example, costly religious rituals are thought to serve the purpose of signaling morality to others (Henrich, 2009; Nesse 2001; Sosis & Alcorta, 2003; Sosis & Bressler, 2003). As another example, people may express disgust (Giner-Sorolla, Kupfer, & Sabo, 2018; Kupfer & Giner-Sorolla, 2017) or anger (Piazza & Landy, 2019) at others’ actions as a signal of their moral values. People apologize when they have committed a transgression to signal reconciliation (Ohbuchi, Kameda, and Agarie, 1989). Political candidates emphatically advertise their trustworthiness and honesty to voters (Funk, 1996). Signaling morality starts at a young age. Even young kids will flip a coin prior to allocating resources to signal to those watching that they are making a fair choice (Shaw, Montinari, Piovesan, Olson, Gino, & Norton, 2014). In essence, people tend to make an effort to make cues about their morality available to those around them.

Judges must then detect these moral cues which are given to them. As previously discussed, people are strongly motivated to seek out moral cues. People preferentially seek out morally relevant information (Wojciszke et al., 1998) because they can use this information to avoid being the victim of immoral acts. And people regularly engage in gossip as a way of distributing morally relevant information among groups (Feinberg, et al., 2012). Indeed, thin slicing research suggests that people are able to detect cues about morality quickly and from simple physical features in a person’s face (Todorov, Said, Engell, & Oosterhof, 2008).

Detecting morally relevant cues will not help people form an accurate impression unless judges utilize moral cues in the appropriate way. For example, misconstruing a sincere apology as
insincere, or misunderstanding a person’s motives could lead a judge to hold a completely inaccurate impression. Much of the work on how people utilize moral cues comes from work that looks at how people ascribe blame to others after an immoral act occurs. In general, this work posits that judges of morally relevant actions work their way through features of the act, such as whether the actor intended the outcome, had control over the outcome, or could foresee the results of the action (Cushman, Young, & Hauser, 2006; Helzer et al., 2017). Based on these features of the act, people form an impression of how much or little blame is warranted. Even actions that are typically rated as less severe transgressions (e.g. hurting an animal vs. hurting a person) can be utilized more when judging a person’s morality when a person is more blameworthy (Tannenbaum, Uhlmann, & Diermeier, 2011).

3 Challenges to forming accurate moral impressions

The process by which people form accurate moral impressions is fraught with challenges, making it easy for judges to end up with inaccurate moral impressions. If something goes wrong at any stage of the impression formation process – a person hiding morally relevant information, a judge missing moral cues, a moral behavior misinterpreted – then a judge will likely end up with an inaccurate moral impression. Below, I describe two big challenges to forming accurate moral impressions.

3.1 What is morally relevant?

Just as impression formation starts from the existence of relevant cues, the first major challenge to forming moral impressions is deciding what is morally relevant. This is a challenge for judges as well as for researchers. First of all, there is no universally agreed categorization of what belongs in the construct of morality and what does not. This has led scholars to define morality in a wide variety of ways (Whiteley, 1960), as evidenced by the term moral character being used differently from literature to literature (Fleeson, Furr, Jayawickreme, Meindl, & Helzer, 2014). Similarly, non-experts vary in what they consider morally relevant. For example, liberals and conservatives place different emphasis on various moral domains (Graham, Haidt, & Nosek, 2009). And open-ended responses to what is morality produce wide varieties of themes and responses between individuals and cultures (Graham et al., 2011). These differences in what is considered morally relevant pose a challenge to forming accurate moral impressions right from the initial relevancy stage of impression formation.
Given that scholars have long striven for and failed to reach a consensus on what defines the moral domain, I focus instead on the traits that are at the core of morality. In other words, there are constructs that are more widely agreed upon as residing in the domain morality than others that can be thought of as central to the moral domain. To get a better idea of which impressions fall squarely in the domain of morality, I looked to scholars who have attempted to map the domains of morality. In Table 1 I present a list of moral domains from several modern scholars. As expected, there clearly is no universal consensus among these scholars on the moral relevance of all constructs. There is, however, some consensus about several domains. The domains of caring (compassion), fairness, honesty, and loyalty all appear on more than half of the lists in Table 1. As such, impressions of compassion, fairness, honesty, and loyalty fall more squarely in the moral domain and can serve as a good starting point for the assessment of moral impressions.

Table 1. Moral Domains According to Various Scholars

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Principled</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dependable</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Integrity</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Caring</strong></td>
<td><em>Caring</em></td>
<td><em>Caring</em></td>
<td>-</td>
<td><em>Caring</em></td>
<td><em>Care</em></td>
</tr>
<tr>
<td>Trustworthy</td>
<td>Trustworthy</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Confident</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Fair</strong></td>
<td>-</td>
<td><em>Fair</em></td>
<td>-</td>
<td>-</td>
<td><em>Fairness</em></td>
</tr>
<tr>
<td>Loyal</td>
<td>Loyal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td><em>Loyalty</em></td>
</tr>
<tr>
<td>-</td>
<td>Respectful</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Respectful</td>
</tr>
<tr>
<td>-</td>
<td>Friendly</td>
<td>Friendly</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>Loving</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>Genuine</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>Honest</td>
<td>Honest</td>
<td>-</td>
<td><em>Honest</em></td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>Generous</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>Hardworking</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>Helpful</td>
<td>Helpful</td>
<td>-</td>
<td><em>Helpful</em></td>
<td><em>Kind</em></td>
</tr>
<tr>
<td>-</td>
<td>Kind</td>
<td>-</td>
<td>-</td>
<td>Kind</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Courage</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Wisdom</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Humanity</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Justice</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Transcendence</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Temperance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Beyond a rough consensus among scholars that compassion, fairness, honesty, and loyalty fall in the moral domain, there are theoretical and empirical reasons for focusing on compassion, fairness, honesty, and loyalty as domains of morality. Theoretically, morality may have evolved to solve a major problem of living in groups or societies. Mainly, groups are often benefited when individuals put their own short-term benefits aside to benefit the group at large (Bloom 2012; Graham & Haidt, 2010). The social-functionalist approach suggests that a common moral code serves the function of binding people into cooperative groups by reducing cheating and free-riding (Fehr & Gachter, 2002; Gurek, Irlenbusch, & Rockenbach, 2006) and promoting pro-social behaviors (Paciello, Fida, Cerniglia, Tramontano, & Cole, 2013). Compassion, fairness, honesty, and loyalty then would fall into the moral domain if they serve similar functional purposes of reducing costly behaviors and promoting pro-social behaviors. Indeed, compassion is by definition the concern for the suffering of others and theoretically should motivate helping behaviors (Batson, 1991; Goetz, Keltner, Simon-Thomas, 2010). Fairness may serve the function of making sure resources are equitably distributed among groups. Honesty can be thought of as a dispositional tendency to refrain from manipulating others for personal gain (Ashton & Lee, 2008). And loyalty is a domain that binds groups together and may serve the function of protecting the group itself (Graham, 2013). In these ways, compassion, fairness, honesty, and loyalty would theoretically fall inside the domain of morality because they serve to promote the welfare of communities.

Empirically, evidence of compassion, fairness, honesty and loyalty as moral domains can be found in at least two sources. First norming data suggest that these domains are rated as highly moral. For example, Goodwin et al. (2014) collected ratings of warmth, morality, and ability on many trait adjectives and honesty, fairness, and loyalty were rated as highly moral and low in warmth. Although they didn’t’ include the trait adjective compassionate, a similar adjective kind was rated as highly moral and warm. This norming data suggest that these domains are not only considered highly relevant to morality, but that they are also relatively uniquely high in morality compared to other-person perception demotions like warmth. Similarly, unpublished data collected as part of piloting studies with the Character Project showed that people consider
compassion, fairness, honesty and loyalty to be highly relevant to morality (Helzer, Furr, & Fleeson, 2013). Second, these domains have been linked to pro-social and ethical behaviors. Compassionate is related to more charitable donations (Ashar et al., 2016) and helping behavior (Valdesolo, P., & DeSteno 2011). Children who are sensitive to fairness violations are more likely to share (Sommerville, Schmidt, Yun, & Burns, 2013). Fairness preferences predict how people behave in negotiations (De Bruyn & Bolton, 2008). Honesty is linked to ethical behaviors in the workplace such as avoiding manipulating others, following rules, and telling the truth (Ashton & Lee, 2008; Lee, Ashton, & de Vries, 2005). Taken together, there is empirical evidence that suggests that these domains are relevant to moral behaviors.

In sum, while there is much more to morality than compassion, fairness, honesty, and loyalty, these domains are likely at the core of morality because they are consistently identified by scholars as part of morality, they are theoretically relevant to morality, and empirically related to moral behaviors. As such, I will focus on impression of compassion, fairness, honesty, and loyalty as indicators of moral impressions.

3.2 Moral traits are not “good” traits

A second big challenge to forming accurate moral impressions arises because the nature of moral traits makes them difficult to judge. Some traits are naturally easier to judge than others. For example, extraversion seems relatively easy to judge, whereas agreeableness tends to be difficult to judge (Connolly, Kavanagh, & Viswesvaran, 2007; Vazire & Carlson, 2010). Indeed, scholars who study accuracy often talk about “good” traits as a moderator to accuracy (Back & Nestler, 2016). A “good” trait is a trait that is neither evaluative nor internal (John & Robins, 1993). From this perspective, moral traits are not “good” traits to judge because they are highly evaluative and moderately internal. Thus, the very nature of moral traits makes them especially difficult to judge.

3.2.1 Evaluative

Moral traits are highly evaluative (Benet-Martínez & Waller, 2002; Goodwin et al., 2014; Helzer, Furr, Hawkins, Barranti, Blackie, & Fleeson, 2014), regardless of how one specifically operationalizes moral traits. There is little to no ambiguity in people’s understanding of what
people view as the “good” or desirable end of the moral spectrum – it is desirable to be seen as moral.

Due to the highly evaluative nature of morality, people are likely to engage in motivated reasoning about their own or other’s morality. For example, when thinking about evaluative domains, people engage in self-presentational motives (Bolino, 1999; Bolino, Kacmar, Turnley, & Gilstrap, 2008; Leary & Kowalski, 1990), are embarrassed to admit socially undesirable things (Fisher, 1993), and activate enhancement motives (John & Robins, 1994; Sedikides & Gregg, 2008). All of these can thus degrade the quality of moral impressions. What is particularly threatening for researchers is the presence of stable individual differences in the tendency to engage in these motivated processes, because these individual differences will be conflated with individual differences in moral impressions. Thus, individual defenses in the tendency to engage in motivated reasoning contribute to systematic variance in moral impressions that clouds true variance due to the underlying moral trait.

While motivated reasoning is mostly thought of as a threat to the quality of self-impressions of morality, it may also be an issue for impressions of other’s morality. In particular, when people are judging a close or liked other, they may engage in the same motivated reasoning processes as they do when rating themselves. As people get closer to others, they tend to include the self in their thinking about the other (Aron, Aron, Tudor, & Nelson, 1991). This merging of the self and other for close relationships may include defense mechanisms and motivated reasoning such as self-presentation or enhancement motives. If so, then informant ratings of a close or liked other’s morality may still suffer a degradation of quality.

3.2.2 Internal

Moral traits are internal (Helzer, et al., 2014). Although some aspects of moral traits such as overt behaviors are easily detectable (e.g., apologies, expressing sympathy, and condescending tone (Bollich et al., 2016), many other aspects of moral traits such as motives and affect reside inside the person. There is evidence that children and adults consider an action as more wrong when harm is the means to an end as opposed to a foreseen side effect (Powell, Derbyshire, & Guttentag, 2012), which requires a judge assessing the internal state of an actor. Similarly, people are judged more favorably after a moral transgression if they exhibit remorse and one of the cues for remorse is guilt expression. Further, people tend to use information about motives
and intentions more when making judgments of morally relevant traits (Kruger & Gilovich, 2004). Likely because the motives or intentions of an actor influence people’s judgments of how blameworthy a person is (Cushman et al., 2006). In other words, people utilize behaviors very differently in the context of information about the internal mental state of actors. An apology can go from a cue of morality as sincere, or a cue of immorality if perceived as insincere. Taken together, this suggests that information internal to the target of a moral judgement might be particularly crucial for moral impression.

3.3 Asymmetries in achieving accuracy

The Self Other Knowledge Asymmetry Model (SOKA) provides an overarching framework for predicting who holds accurate knowledge about a person (Vazire, 2010). In essence, the theory describes several asymmetries between the self and other perspectives. For example, there is an asymmetry in access to information because the self has privileged access to internal thoughts and feelings (Funder, 1999). And there is an asymmetry in motivation because the self is motivated to enhance their self-view (Paulhus & John, 1998, Robins & Beer, 2001, Taylor & Brown, 1988). According to the SOKA Model the consequence of these asymmetries will be two-fold. First, that other-impressions should be more accurate about evaluative traits than the self-impression because the quality of the self-impression will be degraded by self-enhancement motives. In the detection phase of impression formation, other-impression should not be as motivated to ignore self-deprecating cues as selves, and, in the utilization phase, other-impressions are not as motivated to explain away information that might produce a negative impression. Second, self-impression should be more accurate for internal traits than other-impressions. This is primarily because internal cues are readily available to the self, but not to others. Without access to these cues, others aren’t given an opportunity to detect or utilize what is necessary to form an accurate impression.

Although the mechanisms of asymmetry haven’t been directly tested, there has been at least partial empirical support for SOKA Model predictions. For example, there is emerging evidence that others are better judges of highly evaluative traits. When it comes to intelligence, friends have more unique predictive power for IQ test results controlling for the self than the self controlling for friends (Vazire, 2010). When it comes to entitlement, peer reports were the only
unique predictor of social facilitation behaviors at work (Miller & Gallagher, 2016). As such, it is reasonable to think what the SOKA model will predict for moral impression.

Given the highly evaluative nature of morality, the SOKA model predicts that other-impressions should be less biased and more accurate than self-impressions. Others won’t be as motivated to enhance the moral standing of the person they are judging; instead, they may be motivated to form an accurate impression to avoid being the victim of a moral transgression. Yet, given the internal nature of morality, SOKA predicts self-impressions should be more accurate than others. Many of the cues necessary to judge morality are internal (e.g. thoughts, feelings, intentions) and the self has prevailed access to these cues. Taken together, the SOKA model prediction is that the evaluative nature of morality would derail self-perception from becoming accurate, and the internal nature of morality would derail other-perception from becoming accurate. Thus, it may be especially difficult to find any accuracy in moral impressions.

4 Ways of assessing accuracy

There are at least two ways to evaluate if moral impressions are grounded in reality as opposed to idiosyncratically held by the perceiver. The first way is to establish that multiple perceivers agree about the moral impression. Self-other agreement - the tendency for people to see themselves as others see them – establishes that the self agrees with people (or at least someone) around them. Inter-judge consensus – the tendency for multiple others to see a target in the same way – establishes that there is agreement of a shared reputation about a person. The second way to evaluate if moral impressions are accurate is to see if they correspond to observable behavior. This can be done with behaviors observed in the controlled setting of a laboratory or outside of the laboratory as part of people’s everyday lives. The section below outlines the logic, strengths, and pitfalls of utilizing such approaches to evaluating accuracy.

4.1 Self-other agreement and consensus

Agreement between two perceivers has long been used as a proxy for accuracy (e.g. Funder & Colvin, 1991; Funder & West, 1993; Kolar, Funder, & Colvin, 1996; Taft, 1955). If two people hold accurate impressions then they must be in agreement about those impressions (Blackman & Funder, 1998). For example, if in reality Tom is moral, and both Jake and Jess are accurate about Tom’s morality, then Jake and Jess must agree that Tom is moral. With that said, it is still
possible for a perceiver to be in agreement and inaccurate if they hold impressions that are both inaccurate in precisely the same way. As such, the agreement between two perceivers is necessarily true if perceivers are accurate, but not a fully sufficient condition for establishing the accuracy of an impression as grounded truth.

While agreement on moral impression may not be sufficient to establishing accuracy, it is sufficient for establishing a shared social reality. That is, if two people hold the same impression, they share an understanding of what a person is like even if it is wrong. This shared reality may be just as important as accuracy. For example, there is a growing body of research that suggests that people who see their Big Five personality traits more positively than others, do tend to be liked less and lose their social status over time (Back et al., 2011; Carlson & DesJardins, 2015; Colvin, Block, & Funder, 1995; Kurt & Paulhus, 2008; Paulhus, 1998), whereas people who see themselves as others do tend to be liked more in the long run (Human, Biesanz, Parisotto, & Dunn, 2012). Further, failing to share a social reality is diagnostic of mental disorders (Gliga, Senju, Pettinato, Charman, & Johnson 2014; Iwanaga, Tanaka, Nakane, Honda, Imamura, & Ozawa, 2013; Siller, Swanson, & Teachworth, 2014). In sum, indexing the agreement between perceivers is a useful way of assessing a shared social reality, if not reality itself.

4.2 Behavioral Criteria

When two perceivers don’t agree, it can be difficult to determine if either of the perceivers has a realistic moral impression. However, one way to determine how realistic an impression is without relying on shared impressions is to evaluate if an impression is tethered to observable trait-relevant behaviors.

Scholars have designed and employed many in-lab assessments of moral behavior. As examples, in Hartshorne and May's (1928) honesty task, children reported which side of a screen a coin appeared on and fostered conditions whereby students could easily cheat. In the dictator game participants decide how to allocate money between themselves and a partner and there are incentives to make an unfair offer (Bardsley, 2008). In the public goods game participants similarly decide how to allocate resources, but for this task it is between sharing with the group or keeping it for the self (Van Lange, & Liebrand, 1989). In this way, many of the in-lab measures of moral behavior involve pitting the self interest against the benefit of others, or by pitting the self interest against some common social norms of honesty or cheating.
Yet, for moral traits it may be difficult to provide participants an opportunity to behave in relevant ways in the artificial environment of a laboratory. As such, it is ideal to complement in-lab observations with observations of behavior outside the lab. There are many ways to do this. For example, researchers can listen to the ambient noises of participants (i.e. Electronically Activated Recordings; Mehl, Pennebaker, Crow, Dabbs, & Price, 2001) or have participants wear video cameras (Brown, Blake, & Sherman, 2017). But by far the most common way is to ask participants to regularly fill out surveys through a typical day or week (i.e. experience sampling methodology). This can be done once daily (i.e. daily diaries), or several times throughout the day to avoid memory effects.

A few studies have employed these techniques to look at moral behaviors outside of the laboratory and evaluate consistency in behavior. For example, there is evidence that people’s moral behavior is consistent and detectable via audio recordings (Bollich, Doris, Vazire, Raison, Jackson, & Mehl, 2016). This primarily includes interpersonal acts such as showing sympathy or being condescending, because they can more readily be heard on a recording and because grand acts of morality (e.g. running into a burning building) are necessarily rare. These smaller interpersonal acts are the most frequent moral behaviors and account for about 50% of all moral behaviors people reported in their daily lives (Hofmann et al., 2014). Further, there is evidence that moral behavior measured via experience sampling methodologies is consistent from week to week (Meindl, Jayawickreme, Furr, & Fleeson, 2015). Taken together, this suggests that moral behavior can readily be observed in people’s everyday lives and that individual differences in these behaviors would reflect a general tendency to engage in moral behavior. As such, daily reports of moral behavior may serve as a criterion to evaluate the accuracy of moral impressions.
Chapter 2  
Measuring Moral Impressions

In two studies I sought to achieve a number of goals. First, I sought to develop an approach for measuring moral impressions at the core of the moral domain. Importantly, I aimed to develop a modeling approach that quantifies people’s moral impressions while formally accounting for evaluative biases. This is an important first step because evaluative biases systematically confound true variance moral impressions. Thus, models that allow for the measurement of moral impressions without this confounding variance would unlock the possibility of answering poorly or completely unaddressed questions about how well moral impressions reflect a shared reality.

Second, I sought to evaluate if the nature of moral impressions changes with respect to how much a judge likes the person they are rating. Is the nature of morality different for those we liked compared to those that we do not? Perhaps we care more about some moral domains for those that we like, whereas people might care more about others for those that they dislike. For example, people may care more about the loyalty for in-group members when forming a moral impression because a lack of loyalty is the most threatening to the perceiver. Or people may care more about the fairness of out-group members when forming a moral impression, because they stand more to gain from a fair-minded out group member who can put aside group identity. It is particularly important to address this question for theoretical and practical reasons.

The answer to this question is theoretically important for several reasons. First, because if the nature of what constitutes a moral impression fluctuates from judge to judge, and target to target, then moral impressions will necessarily be idiosyncratic and not grounded in a shared reality. Second, because moral impressions are intimately related to liking. Morality plays a central role in how much people like each other (Hartley, Furr, Helzer, Jayawickreme, Velasquez, & Fleeson, 2016).

The answer to this question is also practically important because participants overwhelmingly nominate informants that are liked (Leising, Erbs, & Fritz, 2010), which leads informants to rate people in overly positive ways. If the nature of moral impressions change based on the dynamic of the interpersonal relationship, then most research on morality would be describing a narrower
subset of moral impressions and will not be able to generalize conclusions well. Thus, if the nature of moral impressions is not stable across levels of liking, then researchers would need to use extreme caution when designing multi-method studies and recruiting informants. Similarly, research in moral psychology often asks for participants to rate the morality of a hypothetical person in lab constructed vignettes. If the moral ratings provided by the participants represent different aspects of morality when the characters are likeable or not, then researchers would need to be extremely careful to match the likeability of vignettes across different experimental conditions.

To achieve my first goal of measuring moral impressions while accounting for evaluative bias, I employed a modeling approach that evaluates impressions across a variety of independent evaluative domains (Anusic, Schimmack, Pinkus, & Lockwood, 2009; Kim, Schimmack, & Oishi, 2012). The logic behind this approach is that impressions from evaluative domains may be due to meaningful domain specific variance, systematic evaluate biases, and non-systematic measurement errors. For example, variance in impressions of intelligence are likely the combination of meaningful differences in impressions that are specific to intelligence, meaningful individual differences in how favorably (or unfavorably) evaluate themselves in general, and random measurement error. When an evaluative domain is measured in isolation, the meaningful variance that is specific to a given domain and moral general evaluative biases cannot be differentiated – thus confounding one another. To overcome this, we measured impressions across multiple independent evaluative domains, because independent evaluative domains should share the influence of general evaluative biases but not influences specific to each domain. Thus, by modeling impression in a way that partitions the variance that is shared among items within a single domain and what is shared across all evaluative domains we can purify domain specific variance from more general evaluative biases.

To achieve my second goal of testing the nature of moral impressions at different levels of liking, I turned to the logic of measurement invariance. In structural equation modeling a latent variable is modeled such that it causes observed impressions, but the meaning of a latent variable is tethered to content of what it can predict (Bollen, 2014). As such, in order to be confident that latent variables measured in different groups are capturing the same conceptual content, researchers look to the items that latent variables can predict (i.e., measurement invariance). If a latent variable can predict the same observations in the same way across groups, then researchers
can be confident that the conceptual meaning of the latent variable is somewhat commensurate across those groups. This is sometimes called pattern invariance because the pattern of factor loadings is constrained to equality. Although there are further steps to ensuring latent variables are commensurate (e.g. constrains on variances and intercepts), in practice pattern invariance is the among the important evidence that a latent variable has shared meaning across two groups (Byrne, Shavelson, and Muthén, 1989). As such, we will test if moral impressions of liked, neutral, and disliked others shared the same conceptual meaning by assessing pattern invariance in our moral impressions across these three groups.

As both goals require thoughtful evaluation of structural equation models, it is worth clarifying how I evaluated if a model fits the data well, and how I choose between competing models. In order to determine if my models fit the data well, I will turn to model fit indices. While there are numerous indices of model fit, there is no definitive consensus about which model fit indices and corresponding cut off values are best for testing the fit of structural equation models (Putnick & Borstein, 2016; Cheung & Rensvold, 2002). Indeed, each fit measure has both strengths and some weaknesses. For example, some fit indices are sensitive to sample size, some heavily reward parsimony, some measure goodness of fit, some measure badness of fit, and some have known sampling distributions.

I focused specifically on CFI, RMSEA, and SRMR fit indices because they embody a diverse selection of fit indices. The CFI, a measure of the goodness of fit, is the most widely reported fit index because it is independent of sample size. The RMSEA and SRMR are measures of absolute fit. The RMSEA is regarded one of the most informative fit indices because it takes into account the parsimony of models (Diamantopoulos & Siguaw, 2000). The RMSEA additionally has a known sampling distribution, thus a confident interval can be calculated around this fit estimate, providing a sense of precision around the quality of fit (McQuitty, 2004). The SRMR accounts for sample sizes used to test models and model parsimony. Thus, in order to determine if my models fit the data well, I used established cutoffs of CFI > .95, RMSEA < .08, and SRMR < .08 (Hooper, Coughlan, & Mullen, 2008). Similarly, I will choose between competing models by adhering to the strictest recommendations laid out by Chen (2007), |ΔCFI| < .01, |ΔRMSEA| < .010, and |ΔSRMR| > .03.
1 Study 1 methods

1.1 Participants

Participants were adults from Canada and the United States recruited participants via Amazon’s Mechanical Turk and the undergraduate research pool. There were no exclusion criteria for this study beyond being over 18 years of age and residing in Canada or the United States. Recruitment began via Amazon’s Mechanical Turk by inviting participants to complete a 1-hour online survey about personality in exchange for $10 Amazon gift card. In initially 325 participants were recruited this way; however, participants were only included in analyses if they correctly responded to 9 general knowledge attention check items (e.g., how many moons does earth have?; Full list can be found in Appendix A) resulting in a total of 180 attentive participants.

Given the plethora of inattentive responding, and in order to reduce the financial strain of paid online recruitment participants recruitment continued with an undergraduate research pool for the remainder of the academic year. Undergraduates were invited to take part in a 1-hour survey about personality in exchange for 1-hour undergraduate research participation credit. Initially 642 undergraduates were recruited in this way however, undergraduate participants were only included if they correctly responded to all attention check items resulting in 367 being included in analyses. Thus, a final sample of 547 attentive participants were recruited and included in analyses.

The final sample included participants were typically young adults (M = 26 years, SD = 12 years), but ranged from 18 to 70 years old. The final sample (64% female) was mostly White (49.7%) but was also relatively diverse (6.4% Black; 26.6% Asian; 4.6% Hispanic; 12.7% Other or missing).

No a priori power analyses were conducted prior to data collection to determine sample size because the data was planned to be used to test a variety of research questions beyond this project and each formal test would present differing sample size determinations. Instead, this sample size was the product of ensuring that there were enough participants to get stable estimates, and practical limitations.
The statistical validity of SEM modeling conducted in this work depends on the stability of the underlying correlations among variables. Using the recommendations of Schönbrodt, and Perugini, (2013), and assuming the typical conditions of psychology (correlations of around .21), a sample size of at least 250 would produce stable estimates of correlations (i.e., 80% chance that recruiting additional participants won’t change the estimated correlation by more than .1). The aim was to recruit at least 250 attentive participants by continuing recruitment until the end of the academic semester.

1.2 Procedure

Participants who were recruited on Mechanical Turk completed the survey online, whereas participants who were recruited via the undergraduate research pool completed the survey using a computer in the lab. Although the location differed between Mechanical Turk and undergraduate participants, the procedures and material were otherwise identical.

Study 1 was designed to collect data for a wide variety of research topics. Data for this project come from a subset of those data. Accordingly, the materials and procedures described here represent those which are relevant to the current project. A full list of scales included in Study 1 can be found in Appendix A. All participants completed a computer survey which took approximately one hour to complete. Participants begin the survey by providing demographic information. Subsequently participants rated themselves on a variety of personality traits including moral traits and non-moral evaluative traits. After describing their own personality, participants chose a person from their life that they liked, disliked, neither liked or disliked, and recently met, and then rated them on the same traits as they rated themselves. Participants provided ratings of acquaintances simultaneously in a matrix where columns represented each acquaintance, and rows were personality traits. In total, each participant provided five sets of personality impressions – one about themselves and four about other people in their lives.

1.3 Materials

Table 2 includes a list of traits in the moral domain and non-moral evaluative items.
1.3.1 Moral impression factor

A moral impression factor was modeled using a set of 4 moral impression items. Participants provided impression of compassion towards others, fairness, honesty, and loyalty on a 1 (not at all) to 9 (extremely) Likert scale. Each of these items was selected for their simple face valid representation of different moral domains in order to have a concise set of face valid items that capture a reasonable amount of breadth of the construct of morality. These items were presented in a random order among all personality trait items, including non-moral evaluative items to avoid any order effects.

1.3.2 Evaluative bias factor

An evaluative bias factor was modeled using three non-moral but evaluative items. Participants provided impression of intelligence, physical attractiveness, and funny on a 1 (not at all) to 9 (extremely) Likert scale. Each of these items was selected to represent evaluative but non-moral impressions. These items were presented in a random order among all personality trait items, including moral items to avoid any order effects.

| Table 2. List of Items Used To Measure Moral Impressions and Evaluative Bias |
|---------------------------------|-----------------|-----------------|
| Factor                          | Study 1         | Study 2         |
| Moral Impression                | Honest          | Honest          |
| Moral Impression                | -               | Genuine         |
| Moral Impression                | Loyal           | Loyal           |
| Moral Impression                | -               | Dutiful         |
| Moral Impression                | Fair            | Fair            |
| Moral Impression                | -               | Equitable       |
| Moral Impression                | Compassionate   | Compassionate   |
| Moral Impression                | -               | Caring          |
| Non-moral Evaluative            | Funny           | Funny           |
| Non-moral Evaluative            | Physically Attractive | Physically Attractive |
| Non-moral Evaluative            | Intelligent     | Intelligent     |
| Non-moral Evaluative            | -               | Athletic        |

Note: Items listed together assess the same domain.

2 Study 1 results

2.1 How much evaluative bias is in moral impressions?

All structural equation modeling was conducted using the Lavaan Package in R (version 0.6-1; Rosseel, 2012). I evaluated the influence of evaluative bias on moral impressions in two ways.
First, I modeled the influence of evaluative bias at the latent level by regressing a latent moral impression and an evaluative bias factor. This approach provides a single estimate of how much variance in moral impressions is confounded with evaluative bias and is illustrated in Figure 1a. Second, I modeled the influence of the evaluative factor at the item level by regressing each of the moral impression items on the evaluative bias factor. This approach estimates how much variance in each moral impression item is uniquely the moral domain, an evaluative bias, and measurement error and is illustrated in figure 1b.

As shown in Table 3, moral impressions at the latent level contain a substantial amount of variance due to evaluative bias. When participants reported about their own morality, 26% of the systematic factor variance was attributable to evaluative bias (p < .001). The influence of evaluative bias is even stronger when people rate others’ morality. When participants reported about the morality of other people, estimates ranged from 36.9% when describing a liked other, to 64.8% for neutral others (p’s < .001). Taken together, these results confirm that moral impressions contain statistically significant amounts of evaluative bias.

Furthermore, in Table 3, moral impressions at the item level were partitioned into unique moral variance, evaluative bias, and measurement error. As is typical of all impressions, most of the variance found in single items can be attributed to measurement error (~50%). However, the remaining variance is systematic and can be attributed to either morality or evaluative bias. When participants reported about their own compassion, fairness, honesty, and loyalty, evaluate bias contributed anywhere from 9% for impressions of honesty to 15.1% for impressions of compassion (p’s < .001). However, morality contributed anywhere from 21.6% for impressions of compassion to 38.6% for impressions of loyalty (p’s < .001). When participants reported about others compassion, fairness, honesty, and loyalty, evaluative bias contributed anywhere from 11.4% for impression of compassion for liked others to 42.9% for impressions of honesty for neutral others (p’s < .001).

**Figure 1. Models Used to Partition Evaluative Bias from Moral Impressions**

A)
Note: No equality constraints were imposed on any of these models. Paths not shown were constrained to 0.

Table 3. Quantifying the Relationship Between Evaluative Bias and Moral Impressions at the Factor and Item Levels.
<table>
<thead>
<tr>
<th>Type of Impression</th>
<th>Percentage of Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moral</td>
</tr>
<tr>
<td>Self-impression Factor</td>
<td></td>
</tr>
<tr>
<td>Compassion Item</td>
<td>21.6%***</td>
</tr>
<tr>
<td>Fairness Item</td>
<td>34.8%***</td>
</tr>
<tr>
<td>Honesty Item</td>
<td>37.9%***</td>
</tr>
<tr>
<td>Loyalty Item</td>
<td>38.6%***</td>
</tr>
<tr>
<td>Liked Other-impression</td>
<td></td>
</tr>
<tr>
<td>Compassion Item</td>
<td>33.2%***</td>
</tr>
<tr>
<td>Fairness Item</td>
<td>50.3%***</td>
</tr>
<tr>
<td>Honesty Item</td>
<td>17.1%***</td>
</tr>
<tr>
<td>Loyalty Item</td>
<td>11.0%***</td>
</tr>
<tr>
<td>Neutral Other-impression</td>
<td></td>
</tr>
<tr>
<td>Compassion Item</td>
<td>17.9%***</td>
</tr>
<tr>
<td>Fairness Item</td>
<td>34.5%***</td>
</tr>
<tr>
<td>Honesty Item</td>
<td>16.2%***</td>
</tr>
<tr>
<td>Loyalty Item</td>
<td>07.9%***</td>
</tr>
<tr>
<td>Disliked Other-impression</td>
<td></td>
</tr>
<tr>
<td>Compassion Item</td>
<td>17.9%***</td>
</tr>
<tr>
<td>Fairness Item</td>
<td>27.9%***</td>
</tr>
<tr>
<td>Honesty Item</td>
<td>39.3%***</td>
</tr>
<tr>
<td>Loyalty Item</td>
<td>32.0%***</td>
</tr>
<tr>
<td>New Other-impression</td>
<td></td>
</tr>
<tr>
<td>Compassion Item</td>
<td>20.8%***</td>
</tr>
<tr>
<td>Fairness Item</td>
<td>39.8%***</td>
</tr>
<tr>
<td>Honesty Item</td>
<td>28.4%***</td>
</tr>
<tr>
<td>Loyalty Item</td>
<td>19.3%***</td>
</tr>
</tbody>
</table>

Note: *p < .05, **p < .01, ***p < .001.
2.2 Does liking influence the nature of moral impressions?

In order to test the influence of liking on the nature of moral impressions I tested a series of nested structural equation models that evaluated pattern invariance. In other words, I tested the extent to which moral impressions are similarly anchored to each moral domain. I started with a model that collapsed all evaluative bias factors into a single factor (Figure 2a). However, this model fit poorly (CFI = .826; RMSEA = .086 [.080, .092]; SRMR = .088). This suggests that raters do not employ the same evaluative bias to all targets.

Accordingly, the next model I tested allowed for a separate evaluative bias factor for each target but allowed the bias factors to correlate (Figure 2b). In essence, this model would suggest that raters employ different, but related, evaluative biases for each target person. This model fit well (CFI = .961; RMSEA = .041 [.034, .049]; SRMR = .039) and served as a baseline model to further test a series of nested models.

As shown in Figure 2c, I constrained the factor loadings of the evaluative bias factor to be equal in order to anchor the evaluative bias across domains and groups. This model fit well (CFI = .958; RMSEA = .042 [.035, .049]; SRMR = .040). and imposing these constraints did not meaningfully reduce the model fit (|ΔCFI| < .003, |ΔRMSEA| = .001, and |ΔSRMR| = .001). This suggests that the evaluative bias has the same meaning across different levels of liking.

Then I tested if the influence of evaluative bias on moral impressions is the same across different levels of liking and moral domains (Figure 2d), by imposing equality constraints on all factor loadings from evaluative bias to moral domains to be equal. This model fit well (CFI = .951; RMSEA = .044 [.037, .050]; SRMR = .046) and imposing these constraints did not meaningfully reduce the model fit (|ΔCFI| = .007, |ΔRMSEA| = .002, and |ΔSRMR| = .005). Thus, the influence of evaluative bias is the same for all items and across all levels of liking.

Lastly, I tested for pattern invariance in moral impressions by constraining the factor loadings for each moral domain to be equal across groups (Figure 2e). This model just missed the cut-off for good fit with the CFI but demonstrated good fit with the RMSEA and SRMR (CFI = .947; RMSEA = .045 [.038, .052]; SRMR = .047). Further, imposing these constraints did not meaningfully reduce the model fit for any index (|ΔCFI| = .005, |ΔRMSEA| = .001, and
$|\Delta SRMR| = .002$. Taken together, this suggests that our way of modeling moral impressions carries the same meaning across different levels of liking.

**Figure 2. Models Used to Test for Invariance**

A)

B)

C)
Note: Paths illustrated in black were freely estimated, whereas paths of the same color were constrained to equality. Moral factors were allowed to correlate, evaluative bias factors were allowed to correlate, and residual variances for items from the same moral domain were allowed to correlate. These sources of variance were allowed to correlate from target to target because all of the ratings came from the same people. A person may thus have tendencies to rate morality in similar ways, employ evaluative biases in similar ways, and may get caught up in the language of a specific item in similar ways from target to target. In other words, these sources were all allowed to correlate because there may be perceiver effects regarding morality, evaluative bias, and item specifics. Moral factors, evaluative bias, and item residuals were not allowed to correlate with each other to keep the model identified, and also because I wanted to partition the variance in items into these three unique and independent sources.

3 Study 2 introduction

There are two broad conclusions based on the results from Study 1. First, it is clear that raw moral impressions contain substantial amounts of evaluative bias when people report about the morality of others, and moderate amounts of evaluative bias when people report about their own morality. Second, the nature of moral impressions doesn’t appear to be sensitive to how much a person is liked. Given these two broad conclusions I aimed to extend this work in Study 2. In Study 2, I had three broad goals. First, I sought to replicate our findings from Study 1 regarding the magnitude of influence of evaluative bias on moral impressions using a larger sample.
Second, I sought to get an estimate of moral domain specific variance. Third, I sought to test the convergent validity of the evaluative bias factor.

In study 1, I found that large portions of moral impressions can be attributed to evaluative bias. It is important to replicate the magnitude of influence evaluative bias has on moral impressions because replication is a core component of the scientific method. Although our initial estimates were based on a large and adequately powered sample, researchers have begun to recognize that there are sample specific sources of variance that can contribute to homogeneity in effect size estimates (Judd & Kenny, 2017). For example, there may be effects of the time of year, the scales that collected as part of the broader study, or other study specific methodological choices that influence everyone in the same single large study. Accordingly, we can be more confident in our effect size estimates if we replicate them in a second study that wouldn’t share all of those study specific methods.

In study 1, I employed the logic of Anusic et al., (2009) in order to identify variance in moral impression due to evaluative bias. While the logic behind this approach is sound, it is possible that my label of bias is misleading. Based on Study 1, we can conclude that all evaluative items shared some common cause. However, it is possible – but theoretically unlikely - that a common source of variance in all of the evaluative items is a person being truly higher or lower on all evaluative domains. Indeed, it is difficult to know the nature of a latent factor without anchoring it to specific known measures. In an attempt to empirically justify the label of bias to my evaluative factor, I collected established measures of socially desirable responses to corroborate the bias aspect of the evaluative factor.

In study 1, I found that most of the variance (~50%) in moral domain items couldn’t be accounted for by either a moral impression factor, or evaluative bias, and thus was considered measurement error. However, given that there was only one item per moral domain, any variance that was unique to a given moral domain necessarily would be part of my estimates of measurement error because this variance would not be shared among other moral domains. It is important to estimate moral domain specific variance because past theorizing in moral psychology has suggested there are several distinct but related moral domains (e.g., moral foundations theory) and that these domains are associated with different patterns of thought and behavior. For example, a person may witness someone helping someone else and form an
impression about compassion specifically, but not necessarily morality in general. Thus, my model of moral impressions from Study 1 may be missing out on these domain specific impressions.

In order to replicate and get a more stable estimate of the variance in moral impressions that is due to evaluative bias, I employed the same bi-factor models that I employed in Study 1. I modeled the influence of the evaluative factor at the factor level by regressing a moral impression factor on evaluative bias, and at the item level by regressing each of the moral impression items on the evaluative bias factor.

In order to get an estimate of moral domain specific variance I included an additional item from each of the moral domains. This allowed me to model the variance that is shared among domain similar items that isn’t shared across domains. In essence, including an additional item from each domain allowed me to partition the variance in each item into what is shared across all moral items (broad moral impression), shared among domain similar items (domain variance), shared among all evaluative items (evaluative bias), and what is unique to that specific item (measurement error).

4 Study 2 methods

4.1 Participants

Participants in Study 2 consisted of a sample of adults from Canada and the United States recruited via Amazon’s Mechanical Turk. There were no exclusion criteria for this study beyond being over 18 years of age and residing in Canada or the United States. All participants completed an online survey which took approximately one hour to complete in exchange for a $10 Amazon gift card. The sample (N = 990) included participants were typically middle-aged adults (M = 36 years, SD = 11 years), but ranged from 18 to 72 years old. The sample (51% female) was mostly White (80.2%) with some diversity (10.9% Black; 7.5% Asian; 1.4% Other or missing).

No a priori power analyses were conducted prior to data collection to determine sample size because the data was planned to be used to test a variety of research questions beyond what is relevant here and each formal test would present differing sample size determinations. Instead,
this sample size for this study was a consequence of ensuring that there were enough participants to get stable estimates for an unrelated project that used the same participants.

Just as in Study 1, the recommendations of Schönbrodt, and Perugini, (2013), and assumption of the typical conditions of psychology (correlations of around .21) informed the recruitment of a sample of at least 250 participants in order for modeling to be based on stable estimated correlations (i.e., 80% chance that recruiting additional participants won’t change the estimated correlation by more than .1). Recruitment continued well past 250 total participants in order to meet a sampling quota of participants who mediate for an unrelated study.

### 4.2 Procedure

Study 2 was designed to collect data for a wide variety of research topics in addition to the two goals for this project. Accordingly, the materials and procedures described here represent those which are relevant to the current project’s goals. A full list of scales included in Study 2 can be found in Appendix B.

All participants completed a computer survey which took approximately one hour to complete. Participants begin the survey by providing demographic information. Subsequently participants rated themselves on a variety of personality traits including moral traits and non-moral evaluative traits. After describing their own personality, participants chose a person from their life who knows them well. They were instructed that it could, for example, be a family member, friend, acquaintance, coworker, or romantic partner. Participants then rated them on the same traits as they rated themselves. Unlike study 1, participants rated only themselves and one other target. After rating the personality of a target, participants and completed two measures of socially desirable responding.

### 4.3 Materials

Table 3 includes a list of traits in the moral impression factor items and evaluative bias factor items. A full list of items employed in socially desirable responding scales can be found in Appendix C.
4.3.1 Moral impression factor

A moral impression factor was modeled using a set of 8 moral impression items broken up into 2-item pairs for 4 moral domains. Participants provided impression of the compassion domain by providing impressions of compassion towards others and caring; the fairness domain by providing impressions of fairness and equitable; the honesty domain by providing impression of honesty and genuineness, and the loyalty domain by providing impressions of loyalty and dutifulness. These items were presented in a random order among all personality trait items, including non-moral evaluative items to avoid any order effects.

Additional items which were added to each domain were selected for each moral domain based on discussions in lab meetings which centered around capturing simple synonyms that conceptually match their respective moral domain and remain positively keyed.

4.3.2 Evaluative bias factor

An evaluative bias factor was modeled using four non-moral but evaluative items. Participants provided impression of intelligence, physical attractiveness, funny, and athleticism on a 1 (not at all) to 9 (extremely) Likert scale. Each of these items was selected to represent in dependent evaluative yet non-moral impressions. These items were presented in a random order among all personality trait items, including moral items to avoid any order effects.

4.3.3 Measures of socially desirable responding

I employed two measures of socially desirable responding. The Balanced Inventory of Desirable Responding (BIDR; Hart, Ritchie, Hepper, & Gebauer, 2015) was designed to measure impression management and overly positive responding and was reliable. Participants responded with how much 16 socially desirable but improbably person descriptions applied to them using a 1-7 Likert scale. The BIDR was sufficiently reliable ($\alpha = .85$) and on average participants responded around the midpoint of the scale ($M = 4.3$, 95%CI [4.24, 4.36], $SD = 1.0$, Min $= 1.31$, Max $= 7$). The Social Desirability Scale (SDS; Crowne & Marlowe, 1960) was designed to measure culturally desirable behaviors which are in reality unlikely to occur, such that endorsements of having enacted such behaviors is diagnostic of socially desirable responding. Participants responded to 13 true or false style questions about the desirable behaviors. The SDS was sufficiently reliable ($\alpha = .81$) and participants tended to report that the improbable behaviors
were true and false at roughly equal rates (Min = 0, Max = 13). Both the BIDR, and SDS were presented randomly among all scales in the study.

5 Study 2 results

5.1 How much evaluative bias is in moral impressions?

As shown in Table 4, moral impressions contain significant amounts of evaluative bias. When participants reported about their own morality, 10% of the systematic factor variance was attributable to evaluative bias (p < .001). The influence of evaluative bias is even stronger when people rate others’ morality (36%, p < .001). Taken together, these results confirm that moral impressions contain statistically significant amounts of evaluative bias and provide more precise estimates of the magnitude of variance in moral impressions that is clouded by evaluative bias.

Also, in Table 4 moral impressions at the item level were partitioned into uniquely moral variance, evaluative bias, domain specific variance, and measurement error. As is typical of all impressions, significant portions of the variance found in single items can be attributed to measurement error (~15-50%). However, the remaining variance is systematic and can be attributed to morality, evaluative bias, or domain specific variance. When participants reported about their own compassion, fairness, honesty, and loyalty, evaluative bias contributed anywhere from 4.2% for impressions of loyalty to 7.4% for impressions of compassion (p’s < .001). When participants reported about others’ compassion, fairness, honesty, and loyalty, evaluative bias contributed anywhere from 18.3% for impression of loyalty to 28.4% for impressions of honesty (p’s < .001).

When participants reported about their own compassion, fairness, honesty, and loyalty, a global moral impression contributed anywhere from 35.8% for impressions of a loyalty item to 58.5% for impressions of an honesty item (p’s < .001). When participants reported about others’ compassion, fairness, honesty, and loyalty, a global moral impression contributed anywhere from 33.8% for impression of a compassion item to 48.3% for impressions of another compassion item (p’s < .001).

Table 4. Amount of Variance in Moral Impressions Attributable Morality, Bias, Domain, and Error
<table>
<thead>
<tr>
<th>Type of Impression</th>
<th>Percentage of Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moral</td>
</tr>
<tr>
<td>Self-impression</td>
<td>89.3%***</td>
</tr>
<tr>
<td>Compassion Item 1</td>
<td>40.3%***</td>
</tr>
<tr>
<td>Compassion Item 2</td>
<td>49.8%***</td>
</tr>
<tr>
<td>Fairness Item 1</td>
<td>54.2%***</td>
</tr>
<tr>
<td>Fairness Item 2</td>
<td>42.9%***</td>
</tr>
<tr>
<td>Honesty Item 1</td>
<td>58.5%***</td>
</tr>
<tr>
<td>Honesty Item 2</td>
<td>58.5%***</td>
</tr>
<tr>
<td>Loyalty Item 1</td>
<td>50.4%***</td>
</tr>
<tr>
<td>Loyalty Item 2</td>
<td>35.8%***</td>
</tr>
<tr>
<td>Other-impression</td>
<td>63.5%***</td>
</tr>
<tr>
<td>Compassion Item 1</td>
<td>33.8%***</td>
</tr>
<tr>
<td>Compassion Item 2</td>
<td>48.3%***</td>
</tr>
<tr>
<td>Fairness Item 1</td>
<td>44.6%***</td>
</tr>
<tr>
<td>Fairness Item 2</td>
<td>38.3%***</td>
</tr>
<tr>
<td>Honesty Item 1</td>
<td>43.2%***</td>
</tr>
<tr>
<td>Honesty Item 2</td>
<td>41.7%***</td>
</tr>
<tr>
<td>Loyalty Item 1</td>
<td>47.2%***</td>
</tr>
<tr>
<td>Loyalty Item 2</td>
<td>37.9%***</td>
</tr>
</tbody>
</table>

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

5.2 Estimating domain specific variance

While a global moral impression and evaluative bias accounted for the majority of systematic variance in my items, for some domains and for some impressions there was systematic domain specific variance. When participants were reporting about their own moral impression, only the domain of compassion had systematic domain variance (~25%, p < .001). Whereas when participants were reporting about others, all the domains except loyalty had systematic domain specific variance. The domain of compassion had about 14%, fairness about 5%, and honesty about 4% (p’s < .001). Taken together these results suggest that people largely don’t differentiate between the moral domains when describing themselves (except for compassion), whereas
people do seem to differentiate among the moral domains when reporting about other people (except for loyalty).

5.3 Convergent validity of evaluative bias

As shown in Table 5, I examined the correlations among self and informant reported moral impression and evaluative bias, and scales of socially desirable responding. As expected, there was a very strong correlation among both scales of socially desirable responding \((r = .69, p < .001)\). Similarly, there were strong correlations between self-reports and other reports when it came to moral impressions \((r = .45, p < .001)\) and evaluative bias \((r = .47, p < .001)\). These correlations suggest that people carry with them a tendency to judge morality in similar ways from person to person including themselves, as well as expressing similar evaluative bias from person to person including themselves.

There were significant positive correlations between evaluative bias and socially desirable scales ranging from \(.09\) (other-reported evaluative bias and SDS; \(p < .05\)) to \(.29\) (self-reported and BIDR; \(p < .001\)). On the one hand, these significant correlations are in the predicted direction, which provides some evidence that the evaluative bias factor is measuring socially desirable responding. On the other hand, the magnitude of these effects left much to be desired for drawing any conclusions about the convergent validity of the evaluative bias factor.

Further, there were stronger correlations between socially desirable responding scales and the moral impression factors than there were for evaluative bias factors. An inspection of the items used in each scale, along with these correlations, suggests that although the employed scales purport to measure socially desirable responding, there is likely valid moral variance in these scales. Thus, it is unclear how well these scales serve to establish convergent and divergent validity for my bi-factor model.

<table>
<thead>
<tr>
<th></th>
<th>Self-Reported</th>
<th>Other Reported</th>
<th>Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moral</td>
<td>Evaluative Bias</td>
<td>Moral</td>
</tr>
<tr>
<td>Self-Reported</td>
<td>-</td>
<td>0.00</td>
<td>.45***</td>
</tr>
</tbody>
</table>

Table 5. Correlation Matrix Among Socially Desirable Responding Measures, Moral Factor, and Evaluative Bias
5.4 Are self- and other-reports of morality commensurate?

In order to test if self- and other impressions about morality are commensurate, I employed the same logic as I applied when testing the influence of liking on the nature of moral impressions. I tested a series of nested structural equation models that evaluated pattern invariance. As illustrated in Figure 3, I first modeled self and other impressions free of equality constraints (Panel A). Second, I imposed equality constraints on non-moral evaluative domain items (Panel B; illustrated in purple). Third, I further imposed equality constraints on the influence of evaluative bias on moral domain items (Panel C; illustrated in pink). Fourth, imposed separate equality constraints on the influence of a global moral impression factor on each moral domain item (Panel D; illustrated in gold, blue, green, and red).

Figure 3. Illustration of Nested Models for Self-Other Pattern Invariance

| Evaluative Bias | - | .00 | .47*** | .29*** | .16*** |
| Other Reported |
| Moral | - | 0.00 | .23*** | .17*** |
| Evaluative Bias | - | .15*** | .09* |
| Scales |
| BIDR | - | .69*** |
| SDS |

Note: *$p < .05$, **$p < .01$, ***$p < .001$. 
Note: Paths illustrated in black were freely estimated, whereas paths of the same color were constrained to equality. Moral factors could correlate, evaluative bias factors could correlate, and residual variances for items from the same moral domain could correlate. These sources of variance were allowed to correlate from target to target because all of the ratings came from the same people. Thus, a person may have tendencies to rate morality in similar ways, employ evaluative biases in similar ways, and may get caught up in the language of a specific item in similar ways from target to target. In other words, these sources were all allowed to correlate because there may be perceiver effects regarding morality, evaluative bias, and item specifics. Moral factors, evaluative bias, and item residuals were not allowed to correlate with each other to keep the model identified, and also because I wanted to partition the variance in items into these three unique and independent sources.

As shown in Table 6, my initial model with freely estimated model parameters fit well. Constraining the factor loadings of all non-moral evaluative items on an evaluative factor did not hurt model fit. Nor did constraining factor loadings for all moral domain items on an evaluative factor hurt model fit. Taken together, this suggests that the influence of evaluative bias is commensurate between self and other-impressions. Going further, constraining the factor loadings of each moral domain item to be equal across self and other-impressions did not hurt model fit. This suggests that the influence of a global moral impression on each moral domain
item is the same for self and other-impressions. In other words, the nature of moral impressions appears to be commensurate between the self and other perspectives.

Table 6. Model Fit and ΔModel Fit for Pattern Invariance Models

<table>
<thead>
<tr>
<th></th>
<th>CFI</th>
<th>ΔCFI</th>
<th>RMSEA</th>
<th>ΔRMSEA</th>
<th>SRMR</th>
<th>ΔSRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freely Estimated (Panel A)</td>
<td>.953</td>
<td>-</td>
<td>.055</td>
<td>-</td>
<td>.070</td>
<td>-</td>
</tr>
<tr>
<td>Evaluative Items (Panel B)</td>
<td>.950</td>
<td>.003</td>
<td>.056</td>
<td>.001</td>
<td>.067</td>
<td>.003</td>
</tr>
<tr>
<td>Influence of Evaluative Bias (Panel C)</td>
<td>.946</td>
<td>.004</td>
<td>.056</td>
<td>.000</td>
<td>.077</td>
<td>.000</td>
</tr>
<tr>
<td>Influence of Global Morality (Panel D)</td>
<td>.945</td>
<td>.001</td>
<td>.055</td>
<td>.000</td>
<td>.077</td>
<td>.000</td>
</tr>
</tbody>
</table>

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

6 Study 1 & 2 discussion

Overall, I found that my measure of moral impressions works well. When evaluative variance is partitioned out at the item level, moral impression items contain ample variance that is common to a general moral impression. Further, when multiple items are included for each moral domain, then there is some domain specific variance for other-impressions. Lastly, the nature of moral impressions and evaluative bias in my model are commensurate across other-impressions at different levels of liking and across the self and other perspectives.

My results from Study 1 and 2 strongly suggest that when left unaddressed, evaluative variance confounds meaningful portions of the variance in moral impressions. I found that roughly 10% of moral self-impressions and 30% of moral other-impressions can be attributed to a general evaluative tendency. To put these effect sizes in context we can compare the estimates of how much evaluative bias influences self-impressions of Big Five traits. Evaluative bias in self-reported Big Five impressions tends to account for about 12% of the variance (Anusic et al. 2009). This is very much in line with the amount of evaluative bias I am finding in moral self-impressions.

Much of the theory behind evaluative bias suggests that it should be particularly strong for self-impressions. As examples, the SOKA model predicts that evaluative biases degrade the quality of self-impressions more than other-impressions, and self-enhancement theories predict that
selves are more likely to enhance their standing on desirable traits. Despite this, I found that other-impressions contain far more general evaluative variance than self-impressions. This finding is more in line with work that suggests that people might make more holistic evaluations (good person/bad person) of others instead of evaluating specific domains, whereas we might see ourselves in more nuanced ways.

The different magnitudes of general evaluative variance in impressions suggest that people may judge others on a broad good/bad scale, but maintain more nuance when thinking about themselves. However, the lack of domain specific variance in moral self-impressions compared to other-impressions suggests the opposite. In the moral domain, it appears as though most of the variance in self-impressions that is uniquely moral (i.e., not evaluative bias) is common to an overall moral/immoral impression. Only the domain of compassion showed systematic domain variance, whereas variance in other-impressions had both a strong influence of an overall moral/immoral impression and unique domain specific variance. The domains of compassion, fairness, and honesty all showed significant amounts of systematic domain variance.

Thus, in some ways it seems as though other-impressions are less nuanced because more of the variance in other-impressions can be attributed to an overall evaluation. Yet, in other ways other-impressions show more systematic distinctions between various moral domains. Taken together, this suggests that future work is necessary to describe the kinds of impressions that are more nuanced in self-impressions compared to other-impressions.

There were significant and positive correlations between evaluative bias and socially desirable responding scales; however, the magnitude of these effects was weak to moderate. On the one hand, the lack of strong correlations undermines the convergent validity of the evaluative bias modeling. These scales were developed under the same logic, that states that what is shared among impressions across many evaluative domains can be attributed to an evaluative bias. And the lack of convergence between the scales and the approach I employed undermine such logic. On the other hand, these scales might not have served as the best criteria to validate the evaluative bias. Socially desirable responding scales also were moderately correlated with moral impressions, and an inspection of the items reveals that a majority of the items used in these scales look on the face of them to tap into aspects of the moral domain. Thus, when I scored these scales and treated them as observed manifest variables, they undoubtedly contained a mix
of meaningful variance regarding both socially desirable responding, moral impressions, and other evaluative traits. Future work will be necessary to validate (or invalidate) an interpretation of evaluative bias.
Chapter 3
Measuring Moral Impressions

Equipped with an understanding of how to measure moral impressions, I tested if moral impressions are grounded in reality or if people hold idiosyncratic impressions. Given that people use their moral impressions to inform important life decisions and that moral impressions may be related to how much social value a person has, it is important to know if these consequential impressions are based on some reality or if they are completely in the eye of the beholder.

1 Study 3 methods
1.1 Participants

The total sample consists of undergraduates (N = 465), recruited via the introductory psychology research pool for course credit (2-hours credit per session) or monetary compensation ($20 dollars per in-lab session) and raffle entries for a $50 dollar Amazon gift card for completing daily surveys. Undergraduate participants nominated up to four informants (total N = 481). Informants were compensated with a $10 Amazon gift card. There were no exclusion criteria for undergraduate participants or informants for this study beyond being over 18 years of age. Participants were typically young adults (M = years, SD = years), but ranged from 18 to years old. The sample (75.1% female, 19.6% male, 5.3% unknown) was mostly White (%) with some diversity (% Black; % Asian; % Other or missing).

Only a sub sample of participants completed the daily surveys (N = 184), or returned for session 2 (N = 127) due to attrition. Participants who completed the daily surveys were demographically similar to the full sample (75.0% female, 18.5% male, 6.5% unknown) and did not differ on any of the moral impression factor items or evaluative bias factor items for any t-test comparisons.

1.2 Procedure

Upon deciding to participate in the study, participants came into the lab on two occasions to complete a variety of questionnaires and activities. In between these two in-lab sessions undergraduates filled out daily surveys. This study was designed as part of a larger project and thus contains many questionnaires and activities that are not directly relevant to my work. Below
I’ll describe those procedures that are relevant to my work, but a full list of questionnaires and activities can be found in Appendix D.

In the first in-lab session undergraduates rated themselves on a variety of moral character traits and non-moral evaluative traits and nominated four informants who knew the undergraduates well enough to describe their personality (N = 465). For a week following the first in-lab session, participants completed a week of daily surveys (N = 214). I emailed participants links to each daily survey at 12, 3, 6, and 9pm daily. In the second in-lab session undergraduates came into the lab in groups of four to six and played six the public goods game together (N = 127). Participants were scheduled for the second session such that they would not overlap with participants who they may have met as part of session 1.

I emailed informants and invited them to participate in the study. Those who accepted the invitation subsequently completed several measures about themselves and about the personality of the undergraduate who nominated them as part of a larger study. These measures took roughly an hour to complete and a full list can be found in Appendix D. Of primary interest to this study informants rated undergraduates on a variety of moral character traits and non-moral evaluative traits.

I did not collect identifying information that would be useful in differentiating the informants of each participant. Accordingly, each of the three informants per participant are exchangeable (i.e. we could interchange any one for another). Due to the exchangeability of informants there should not be any systematic or theoretical reasons why the relationship between any two informants will be different from any other two (Kenny, Kashy, & Cook, 2006). In other words, only random chance would account for fluctuations in the strength of relationship between informants. Thus, I followed path analytic procedures outlined in Furr and Wood (2013), whereby parameters that link informants will be constrained to equality. This has the added benefit of being a more parsimoniously model with a single parameter to quantify inter-judge consensus.

No a priori power analyses were conducted prior to data collection to determine sample size, instead recruitment was again designed to recruit enough participants to get stable estimates of correlations. Following recommendations of Schönbrodt, and Perugini, (2013), and assumption of the typical conditions of psychology (correlations of around .21) I aimed to recruit 250
participants with data from session 1, daily surveys, and session 2. Unfortunately, recruitment fell short of the 250-participant goal and was ended after 4 years due to practical constraints such as funding and access to laboratory space. No attention checks were included as part of this study, and no participants were excluded from analyses other than due to missing data.

1.3 Materials

1.3.1 Moral impression factor

A moral impression factor was modeled using a set of 4 moral impression items. Participants provided impression of compassion towards others, fairness, honesty, and loyalty on a 1 (not at all) to 9 (extremely) Likert scale. Each of these items was selected to match with Study 1 and were presented in a random order among all personality trait items, including non-moral evaluative items to avoid any order effects.

1.3.2 Evaluative bias factor

An evaluative bias factor was modeled using three non-moral but evaluative items. Participants provided impression of intelligence, physical attractiveness, and funny on a 1 (not at all) to 9 (extremely) Likert scale. Each of these items was selected to represent independent evaluative but non-moral impressions matching Study 1. These items were presented in a random order among all personality trait items, including moral items to avoid any order effects.

1.3.3 Public goods game

Participants engaged in six rounds of a typical public goods game. In rounds participants decided how many of 20 tokens to contribute to a community chest or keep in their private collection. Contributions to the community chest would be multiplied by 2, and then divided evenly among everyone in the session. As a group, earnings are maximized when more tickets were contributed to the community chest. In contrast, an individual’s earnings were maximized by withholding contributions to the community chest. Thus, contributions to the community chest can be thought of as cooperating for the benefit of the group as a whole at a small cost to the individual. I averaged participants contributions to the community chest across all six rounds to create an index of how cooperative each participant was during the public goods game.
1.4 Daily behaviors

I prompted participants to “think about what you were like right before receiving our notification. Just prior to receiving our notification, to what extent were you thinking, feeling, and/or behaving in the following ways?”. Participants reported about the extent to which they were cooperative, generous, stood up for someone, shared resources or responsibilities (e.g. time, money, food, chores), and provided help or offered help without being asked on a 1 (not at all) to 9 (extremely) Likert scale. These behaviors were chosen based on laboratory discussions in which we tried to identify behaviors that might occur reasonably frequently enough to be captured in a typical week. The aim was to strike a balance between behaviors that were neither too specific as to not be relevant to many participants in a given week, nor too broad as to render the behaviors meaningless while trying to capture behaviors that represent different moral domains. Similarly, these behaviors were designed to map onto domains participants reported on using self-reported moral impression items. For example, sharing resources was selected to fall within the fairness domain. However, an emphasis was placed on behaviors in the fairness domain in order to match natural behaviors to the moral domain theoretically evaluated by the public goods game. Additionally, extreme moral behaviors (e.g., running into a burning building to save someone stuck inside) were avoided which were unlikely to occur in a given week for university students.

These behaviors were all significantly positively related to each other, even after accounting for multiple comparisons (.32 < rs < .69, ps < .001). Thus, I created a composite scale by averaging these behaviors together. This scale showed excellent reliability (α = .86) and participants tended to respond around the midpoint (M = 4.5, SD = 1.42, Min = 1.12, Max = 8.45). In total I collected 2071 daily reports from participants from 214 unique participants (46% response rate). Those participants who completed the daily reports completed an average of 9.6 reports over the course of a week. All participants were invited to take part in the daily surveys, and any participant that completed at least 1 daily report of behavior was included in analyses.
2 Study 3 results

2.1 Inter-judge consensus

As illustrated in Figure 4, I tested for inter-judge consensus by setting up a model that controls for evaluative bias for informants and constrained factor loadings to be equal for each informant. This model had mediocre fit as fit indices fell outside my cut-off values for the CFI and SRMR (CFI = .902, RMSEA = .074 95%CI [.051, .095], SRMR = .095). There was little inter-judge consensus across moral and non-moral impressions. Informants did not agree on a general moral impression of participants ($r = .17, p = .296$). Similarly, informants did not agree about most domain specific impressions of loyalty ($r = .03, p = .758$), fairness ($r = .14, p = .442$), compassion ($r = .20, p = .191$), or intelligence ($r = .26, p = .144$). However, informants did form some consensus about the domain of honesty ($r = .39, p < .001$), physical attractiveness ($r = .28 p = .006$), and funny ($r = .34, p = .001$). And informants formed some consensus about a general evaluative impression of participants ($r = .25, p = .027$).

**Figure 4. Model of inter-judge consensus**
Note: \(* p < .05, \ ** p < .01\). All estimates presented are standardized and were significant at the \( p < .001 \). The correlations between latent variables and item residuals represent the magnitude of interjudge consensus. Paths that are inked in black were freely estimated, whereas paths inked in the same color were constrained to be equal. Yellow, Blue, Green, and Red paths represent the factor loadings of moral impressions on each moral domain item. Pink paths represent the influence of evaluative bias on each moral domain item. Purple paths represent the influence of evaluative bias on each non-moral evaluative domain item.

### 2.2 Self-other agreement

Given that my model of interpersonal consensus did not fit particularly well, and that estimates were based on a limited (\( n = 128 \)) observations, I decided to average informant impressions together with the idea of taking advantage of improved validity of what is shared among
informants (Zou, Schimmack, & Gere, 2013). In doing so, I could base my estimate of agreement on more pairs of self-other observations.

As illustrated in Figure 5, I tested for self-other agreement by setting up a model that controls for evaluative bias but allows for factor loadings to be freely estimated. This model had good fit as fit indices fell well within my cut-off values for the CFI, RMSEA, and SRMR (CFI = .965, RMSEA = .038 95%CI [.011, .058], SRMR = .05). There was agreement between self and other impressions across moral and non-moral impressions. There also was agreement about general moral impressions of participants ($r = .23, p = .011$). Similarly, informants and participants agreed about domain specific impressions of loyalty ($r = .17, p = .020$) and compassion ($r = .29, p < .001$) and non-moral evaluative domains of intelligence ($r = .23, p = .019$), physical attractiveness ($r = .15 p = .048$), and funny ($r = .29, p < .001$). However, informants and participants did not agree about general evaluative impressions of participants ($r = -0.07, p = .497$). Similarly, informants and participants did not agree about the moral domains of honesty ($r = .02, p < .737$) or fairness ($r = -0.06, p = .767$). Taken together, these effects suggest that to some extent people share a social reality about their morality.

2.3 Links between moral impressions and behavior

As illustrated in Figure 6, I evaluated the relationships between in-lab behavior and daily behavior separately. These relationships were modeled separately because many participants failed to either complete the in-lab task or the daily surveys, which prevented me from modeling the relationship between the two behavioral measures. However, I could recapture participants that completed one or the other behavior measures by modeling them separately. This improved the stability of estimates in both models; however, estimates are still based on relatively small sample sizes.

In Panel A of Figure 6, I examined the relationship between self- and other-impression with an in-lab behavior of the public goods game. Self-impressions of morality were uncorrelated with contributions in the public goods game. However, there was a positive correlation between other-impressions of morality and contributions in the public goods game, such that those participants that were rated as more moral by their informants contributed more to the public good. This suggests that others hold realistic impressions of a person’s morality, but not the self.
Figure 5. Structural Equation Model of Self-other Agreement

Note. *p < .05, **p < .01. All estimates presented are standardized and based on n = 242 observations. All factor loadings were statistically significant at the p < .01 level with the exception of self-reported fairness (p = .014) and compassion (p = .088) on self-reported evaluative bias factor.

In Panel B of Figure 6, I examined the relationship between self- and other-impressions with daily ESM reports of moral behavior. Neither self-impressions nor other-impressions of morality were correlated with daily reports of moral behavior. This suggests that neither impressions are tethered to daily behaviors in a way that would make them realistic.
Taken together, there is little evidence that moral impressions are tethered to measures of moral behavior. However, this conclusion should be considered preliminary as my estimates are based on a relatively small sample of observations given the complexity of the models.

**Figure 6. Relationships Among Behaviors, Self-, and Other-reports of Morality**

A) 

B) 

*Note.* *p* < .05, **p** < .01. All estimates presented are standardized. Due to missing data, *n* = 65 for Panel A estimates of correlations with the public goods game, and *n* = 104 for Panel B.
estimates of correlations with ESM reports. Model parameters not shown in the figures are paths previously estimated in Figure 5 with a much larger sample thus providing more stable estimates of those parameters.

3 Study 3 Discussion

The existence and magnitude of a shared reality among perceivers has many theoretically interesting implications for scientists. Indeed, one of the main turning points in the debate over whether personality existed (i.e., person/situation debate) was the robust finding that people often reach consensus when judging an individual’s personality (Kenny & West, 2010), and those judgments often agreed with the target’s self-perceptions (Connolly et al., 2007; Vazire & Carlson, 2010). In order to assess the extent to which moral impressions are shared and grounded in reality, I assessed inter-judge consensus, self-other agreement, and convergence with in-lab and daily behaviors.

In regard to inter-judge consensus, I did not find that two judges formed similar impressions of a shared target. This may be because in reality there is no consensus among judges about how moral a person is. However, there is some past work that suggests that there is consensus among judges (Helzer et al., 2014). It becomes difficult to interpret this null finding in the context of a relatively underpowered empirical test. I might not have had a large enough sample to detect a consensus among judges. Future research will need to recruit larger samples of independent judges in order to better estimate the magnitude of consensus (if any) among other-impressions.

Although consensus was lacking among other-impressions, there was significant agreement between self-impressions and other-impressions. This agreement suggests that impressions of morality are neither completely idiosyncratic to the perceiver, nor completely shared among perceivers. Instead, there is small to moderate convergence between perceivers ($r = .23$). To help interpret this effect size we can employ the logic behind the binomial effect size display (Rosenthal & Rubin, 1982). If a person thinks of themself as above average in morality, there is a 61.5% chance that others also see them as above average in morality.

Self and other-impressions agree to a certain extent, but only other-impressions showed signs of being tethered to morally relevant behavior. Other-impressions were weakly linked to an in-lab
cooperation behavior, but not self-impressions. Worse yet, neither self nor other-impressions were related to daily behaviors.

Taken together, these effects suggest that moral impressions are neither completely idiosyncratic, nor fully tethered to a shared social reality or behavior. These effects suggest that a shared social reality might emerge between, but that this shared impression may be a shared illusion. Or at least that these shared impressions are not usefully predictive of people’s behavior in any given moment.
Chapter 4
What are the Consequences of Sharing Moral Impressions?

Given that moral impressions may be related to social benefits and my estimates suggest that these impressions are only somewhat shared among perceivers, I ask whether there are benefits of sharing a social reality? Do those who share an understanding of their moral character with others benefit from this shared understanding? It may be the case that social benefits of moral judgments are bolstered by the degree to which people’s self-impressions of moral character align with other people’s impressions of their character. That is, sharing a social reality regarding moral character may confer benefits above and beyond those conferred by being seen as moral or seeing oneself as moral.

There is modest to strong agreement that the Big Five (Connolly et al., 2007) agreement for moral impressions is far from perfect ($r = .23$), which raises interesting questions about how practical or advantageous accuracy is. On the one hand, people who argue for the adaptiveness of positive illusions over self-knowledge would argue that blind spots are likely in the direction of positive bias and are there for good reason. That is, disagreements likely stem from self-views being higher than others’ views or other criteria, and people who are positively biased experience better outcomes. On the other hand, people who argue for the adaptiveness of self-knowledge would argue that while there is variability in who has or lacks self-knowledge, people with self-views that are tethered to reality are more adjusted than people who are deluded. Indeed, people who see themselves as others do are good leaders (Sosik & Megerian, 1999). The current section explicitly tests if agreement of moral character is adaptive or not.

1. Is a shared reality adaptive?

The idea that sharing a social reality might confer benefits or that failing to share a reality may be detrimental has been gaining support in other domains. For example, there is a growing body of research that suggests that people who see their Big Five personality traits more positively than others do, tend to be liked less and lose their social status over time (Back et al., 2011; Carlson & DesJardins, 2015; Colvin et al., 1995; Kurt & Paulhus, 2008; Paulhus, 1998), whereas people who see themselves as others do, tend to be liked more in the long run (Human et al., 2012). As such, a similar pattern might be observed for moral impressions.
People who are seen by others and who see themselves as high in moral character are likely to experience more positive outcomes than do people who are seen by others and who see themselves as lacking moral character. However, we argue that there are costs associated with failing to see eye to eye with others regardless of people’s standing on character. Take the example of Jane and Tom. Jane will like Tom if she thinks he is moral (Goodwin et al., 2014; Hartley et al., 2016; Wojciszke, 2005; Wojciszke, Bazinska, & Jaworski, 1998), but Jane’s positive evaluation of Tom might be attenuated if he sees himself as more moral than she sees him. Tom might appear to be arrogant if he claims to be more honest than Jane thinks he is, or their differing viewpoints might create interpersonal conflict (e.g. Jane might get annoyed by Tom’s constant moral bragging). On the other end of the spectrum, while Jane will likely see Tom in a less positive way if she thinks he is immoral, she might like him more if he shares her negative impression, essentially admitting he is immoral, than if he sees himself more positively than she sees him. This prediction is based on a growing body of work that suggests that people enjoy individuals who acknowledge their flaws more than people who do not (Ward & Brenner, 2006) and that people who are aware of having a fairly undesirable reputation are enjoyed more than are people who are less aware of their undesirable reputation (Carlson, 2016; Oltmanns, Gleason, Klonsky, & Turkheimer, 2005). In sum, we suggest that disagreement about moral impressions is maladaptive (i.e. it could attenuate the positive interpersonal outcomes associated with making a good moral impression) whereas agreement is adaptive (i.e. it could attenuate the negative interpersonal outcomes associated with making a bad moral impression). To test this general hypothesis, the current research measures the interpersonal consequences associated with seeing or failing to see eye to eye with others about one’s moral character.

2 Using RSA to test associations with shared reality

While past work has examined outcomes associated with (dis)agreement of personality judgments, traditional approaches to indexing agreement fail to capture the more nuanced relationships between agreement and outcomes (Barranti, Carlson, Cote, 2017). For example, some work looks at difference scores; however, difference scores are plagued with unreliability (Edwards 2002; Shanock, Baran, Gentry, Pattison, & Heggestad, 2010). Some look at profiles; however, this approach conceptualizes agreement in a somewhat unintuitive way (i.e., a characteristic pattern of traits) and estimates can be difficult to interpret once normative and distinctive sources of agreement are disentangled. Some work employs the Truth and Bias;
however, this modeling approach arbitrarily deems one perspective the truth and cannot formally compare different kinds of matches. Thus, past work has been unable to properly test if it is adaptive to share an understanding of your own moral character with others.

To index the link between disagreement and social value, we employ Response Surface Analysis (RSA), an approach based on polynomial regression that models the ways in which all possible combinations of self- and other-impressions predict social value. We believe this approach is superior to other methods for three reasons. First, unlike difference scores, which can have problems with discriminant validity (Furr, 2011) and which impose potentially unwarranted constraints on parameters (Edwards, 1994), RSA retains information about the main effects of self- and other-impressions when modeling how discrepancies predict social value (Edwards, 1994; Edwards, 2002; Cafri, van den Berg, & Brannick, 2010). Second, unlike other more conventional approaches (e.g. residual scores, moderated multiple regression), RSA provides a visualization of these effects in three-dimensional space. That is, rather than relying on point estimates to visualize the data (e.g. +/- 1SD in moderated multiple regression), RSA reveals the full spectrum of how combinations of self-and other-impressions of moral character predict social value. Third and perhaps most importantly, RSA provides statistical tests that directly evaluate, which allows me to test novel but theoretically central questions about whether agreement is adaptive.

This general hypothesis may manifest itself in several ways. First, in its most pure form, people may have optimal levels of adaptive constructs when they see themselves in the same way as others see them at all levels of morality. This prediction, which we call the pure discrepancy hypothesis, is not unreasonable given that self-other disagreement seems to be linked to negative social outcomes in other domains. Second, there may be detrimental effects of disagreement in general, but these effects may compete with benefits of humility. This prediction, which we call the optimal humility hypothesis, suggests that optimal levels of an adaptive construct are associated with some, but not too much humility.
3 Study 4 methods

3.1 Participants

A community sample of 100 “judges” (60% female) were recruited via advertisements on popular websites, in local newspapers. Judges provided the names and email addresses of up to six “targets” (47.5% friends, 24.9% family, 11.0% coworkers, 6.6% partners, 2.4% acquaintances, 7.7% other or unknown) they knew well ($M = 11.94$ years, $SD = 11.84$ years). Judges were eligible to participate if at least five of their nominated targets participated. Recruitment continued until a sample of 100 judges was collected. As part of a larger online study, judges received up to $80$ in Amazon.com gift cards and responding targets ($N = 587$; 55.7% female; $M_{age} = 34$ years, $SD = 13.5$ years) received up to $25$ in Amazon.com gift cards for their participation.

No a priori power analyses were conducted prior to data collection to determine sample size because the data was planned to be used to test a variety of research questions beyond what is relevant here and each formal test would present differing sample size determinations. Recruitment of 100 “judges” each with 6 “targets” was made to maximize power for unrelated questions about person differential accuracy for Big 5 impressions. No attention checks were included as part of this study, and no participants were excluded from analyses.

3.2 Materials and procedure

Judges rated each of their targets’ broad moral character, moral character traits, as well as their extraversion and agreeableness, and targets described themselves using the same measures re-worded as self-reports. Self- and other-impressions of broad moral character included the following three items rated a 1 (Strongly Disagree) to 5 (Strongly Agree) scale: “I am (person X is) a person of strong moral character”, “I would say that I am (person X is) a good person”, and “I (person X) usually do (does) the right thing, even if it's hard” (other-impressions: $M = 4.14$, $SD = .70$, $\alpha = .81$; self-impressions: $M = 4.10$, $SD = .57$, $\alpha = .70$). Self- and other-impressions of each of the four moral character traits (i.e. compassion, honesty, loyalty, fairness) were measured with 3 items on a 1 (Strongly Disagree) to 5 (Strongly Agree) scale. Example items included "I am (person X) a compassionate person" (compassion: other-impressions: $M = 4.11$, $SD = .74$, $\alpha = .86$; self-impressions $M = 4.19$, $SD = .67$, $\alpha = .79$), "I (person X) always tell the truth" (honesty:
other-impressions: $M = 3.86, SD = .78, \alpha = .86$; self-impressions: $M = 3.88, SD = .74, \alpha = .81$), "I do (person X) not shift my (his/her) loyalties easily", (loyalty: other-impressions: $M = 4.14, SD = .66, \alpha = .80$; self-impressions: $M = 4.06, SD = .62, \alpha = .63$), and "I treat (person X treats) people fairly" (fairness: other-impressions: $M = 4.10, SD = .64, \alpha = .79$; self-impressions: $M = 4.14, SD = .57, \alpha = .60$). Self and other-impressions of extraversion and agreeableness were measured using 4-item scales, with an item from each of the four facets of the extraversion (X) and agreeableness (A) factors of the HEXACO scale (Lee & Ashton, 2004) and rated on a 1 (Strongly Disagree) to 5 (Strongly Agree) scale. Example items were "In social situations, I'm (person X is) usually the one who makes the first move" (extraversion: other-impressions: $M = 3.50, SD = .71, \alpha = .62$; self-impressions: $M = 3.41, SD = .74, \alpha = .65$) and “I (person X) generally accept(s) people’s faults without complaining about them" (agreeableness other-impressions: $M = 3.12, SD = .83, \alpha = .74$; self-impressions: $M = 3.00, SD = .71, \alpha = .57$).

Judges rated how much they liked (e.g. “I like person X”, $M = 4.47, SD = .66, \alpha = .84$) and respected (e.g. “I respect person X”, $M = 4.19, SD = .81, \alpha = .81$) each target using 2-items, 1 (Strongly Disagree) to 5 (Strongly Agree) scales.

### 3.3 Analytic approach

RSA effects were modeled in R with the RSA package (Schönbrodt, 2015). RSA uses coefficients from polynomial regression to plot how all possible combinations of impressions predict social value in a three-dimensional space and provides four coefficients ($a_1$ – $a_4$) that statistically test how these combinations predict social value. Polynomial regression coefficients are derived from a multilevel polynomial regression to account for the nesting of other-perceptions within targets. Rather than directly interpreting the five polynomial slopes, I focus my attention on the four RSA coefficients to better understand if and how self-other disagreement about morality is associated with social value.

The $a_1$ coefficient tests the linear slope along the line of perfect congruence ($Y = X$ axis) and reveals if matching self- and other-impressions at higher versus lower levels of moral character is associated with higher or lower social value. Given that moral character impressions are socially desirable (Goodwin et al., 2014; Helzer et al., 2014), the $a_1$ should be positive, suggesting that social value is higher when self- and other-impressions match at higher levels of moral character and lower when matched at lower levels of moral character. The $a_2$ coefficient tests the curvature
of the line of perfect agreement and reveals if matches at low or high moral character (i.e., matches at extremes) are different than matches at moderate levels of moral character. However, I do not anticipate significant $a_2$ coefficients.

My main focus is on the pattern of $a_3$ and $a_4$ coefficients, which test effects are along the line of incongruence ($Y = -X$ axis). The $a_4$ coefficient tests the curvature of the line of incongruence such that the direction (positive or negative) and significance of the $a_4$ indicates whether the magnitude of discrepancy between self- and other-impressions predicts liking or respect. The $a_3$ coefficient tests the slope of the line of incongruence. The direction and significance of the $a_3$ indicates if the direction of disagreement predicts liking and respect. Evidence for the Pure Discrepancy Hypothesis will be observed if there is a negative $a_4$ coefficient (i.e. concave curvature) in isolation with no $a_3$ coefficient, suggesting that as the magnitude of discrepancy between self- and other-impressions of moral character increases, social value decreases. In contrast to our hypothesis, a positive $a_4$ coefficient would suggest a convex curvature, meaning that as the magnitude of discrepancy increases, social value increases. Evidence for the Optimal Humility Hypothesis will be observed if there is both a negative $a_4$ and a positive $a_3$ coefficient. This pattern of effects would suggest that people who see themselves as somewhat less moral than their close other sees them are liked or respected the most. In contrast to this hypothesis, negative $a_3$ and $a_4$ coefficients would suggest that people who see themselves as somewhat more moral than their close other sees them are liked or respected the most. Figure 7 illustrates the shape of the response surface I should observe if each of our predictions are supported.

Figure 7. Visualization of Response Surfaces That Would Support Hypotheses

A) Null Hypothesis
3.3.1 Controlling for evaluative bias

It is unlikely that evaluative bias would drive any effects of agreement because evaluative biases aren’t shared among self and other-impressions. However, given that upwards of a third of systematic variance in moral impressions may be due to these evaluative biases, it is worth ensuring that evaluative bias did not drive effects of agreement. Response surface analyses, while the most appropriate formally test of my hypotheses, has not yet been adapted to be used alongside structural equation modeling in a way that would allow me to model evaluative bias in the same way as it was modeled in Studies 1-3. Further, I did not collect data on the independent evaluative domains used to model evaluative bias in Study 4, which would have allowed me to enter them as a control variable. In order to overcome this limitation and to ensure that evaluative bias was not driving our results, we controlled for extraversion and agreeableness for
all analyses. Agreeableness and extraversion are similarly influenced by participants’ evaluative biases, thus any unique effects of our moral impressions controlling for agreeableness and extraversion should be relatively free form evaluative biases.

4 Study 4 results

My sample included judge-target pairs that varied in the magnitude and direction of discrepancies. As suggested by Fleenor, McCauley, & Brutus (1996) and Shanock et al. (2010), I approximated agreement and discrepancy by considering judge-target pairs whose self- and other-impressions were within half a standard deviation of each other after standardization as roughly in agreement and outside of that as discrepant. About 50% of targets agreed with their judge (broad moral character 49.9%; compassion 46.7%; honesty 46.2%; loyalty 46.3%; fairness 50.1%), whereas 25% of targets saw themselves as having more moral character such that self-impressions were higher than other-impressions (broad moral character 23.7%; compassion 29.5%; honesty 27.7%; loyalty 23.7%; fairness 26.1%). Similarly, 25% saw themselves as having less moral character than their judge saw them as such that self-impressions were less than other-impressions (broad moral character 26.4%; compassion 23.8%; honesty 26.1%; loyalty 30.0%; fairness 23.8%). Thus, my sample contains information about the various combinations of self- and other-impressions of morality necessary to test my hypotheses.

With a sample that contained a diverse set of combinations of self- and other-impressions, I ran polynomial regressions using multilevel modeling to account for the nesting of judges within targets. I then used the polynomial slopes to conduct RSA. In general, I observed evidence for the Optimal Humility Hypothesis in moral impressions. There were consistently negative a4 and positive a3 coefficients for moral impressions. The liking of participants was maximized when self-impressions were humble with respect to other-impressions of compassion, honesty, and loyalty. Similarly, the respect of participants was maximized when self-impressions were humble with respect to other-impressions of a general moral impression, compassion, honesty, and loyalty. This pattern of effects would suggest that people who see themselves as somewhat less moral than their close other sees them, are liked or respected the most.

While the pattern in general supported the Optimal Humility Hypothesis, there was some nuance among the moral domains. In particular, for impressions of fairness the evidence did not support the Optimal Humility Hypothesis but instead the pattern of RSA coefficients and polynomial
regression slopes suggested that only other-impressions predicted liking and respect. Those participants who were seen as fair tended to be liked and respected, regardless of how they saw themselves (liking: $\beta = .65$, $p < .001$; respect: $\beta = .58$, $p < .001$). This same pattern appeared for broad moral character impression when predicting liking ($\beta = .61$, $p < .001$), and for impressions of loyalty when predicting respect ($\beta = .51$, $p < .001$). Taken together this suggests that in some contexts other-impressions may be particularly important and overpower any effects of agreement or self-impressions when it comes to social value.
Table 7. Polynomial Slopes and Response Surface Coefficients for Moral Impressions, Extraversion, and Agreeableness [95% CI]

<table>
<thead>
<tr>
<th>Impression</th>
<th>Outcome</th>
<th>MLM Slopes</th>
<th>RSA Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β₀</td>
<td>β_target</td>
<td>β_target²</td>
</tr>
<tr>
<td>Broad Moral Character</td>
<td>Liking</td>
<td>3.81***</td>
<td>0.61***</td>
</tr>
<tr>
<td></td>
<td>Respect</td>
<td>3.58***</td>
<td>0.67***</td>
</tr>
<tr>
<td>Compassion</td>
<td>Liking</td>
<td>4.05***</td>
<td>0.50***</td>
</tr>
<tr>
<td></td>
<td>Respect</td>
<td>3.76***</td>
<td>0.52***</td>
</tr>
<tr>
<td>Honesty</td>
<td>Liking</td>
<td>4.25***</td>
<td>0.29***</td>
</tr>
<tr>
<td></td>
<td>Respect</td>
<td>3.86***</td>
<td>0.47***</td>
</tr>
<tr>
<td>Loyalty</td>
<td>Liking</td>
<td>4.04***</td>
<td>0.44***</td>
</tr>
<tr>
<td></td>
<td>Respect</td>
<td>3.68***</td>
<td>0.51***</td>
</tr>
<tr>
<td>Fairness</td>
<td>Liking</td>
<td>3.75***</td>
<td>0.64***</td>
</tr>
<tr>
<td></td>
<td>Respect</td>
<td>3.49***</td>
<td>0.58***</td>
</tr>
<tr>
<td>Extraversion</td>
<td>Liking</td>
<td>3.81***</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Respect</td>
<td>3.58***</td>
<td>0.03</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>Liking</td>
<td>3.81***</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Respect</td>
<td>3.58***</td>
<td>0.06*</td>
</tr>
</tbody>
</table>

Note. CI = confidence interval. When estimating extraversion and agreeableness coefficients we used our measure of broad moral character to control for moral impressions. Statistically significant effects are in bold. *** p < .001, ** p < .01, * p < .05, *p < .10. Table reproduced from (Barranti, Carlson, & Furr, 2016).
Figure 8. Response Surfaces of Moral Impressions, Extraversion, and Agreeableness
Note: Lines of perfect agreement (Y = X axes) are illustrated as solid lines and lines of incongruence (Y = -X axes) are illustrated as dashed lines. Reproduced from (Barranti, et al., 2016).

We similarly tested our hypotheses for both extraversion and agreeableness controlling for broad moral character. For extraversion and agreeableness, the results were mostly consistent with the null hypothesis. The pattern of effects suggested disagreement between self- and other-perspectives were completely unrelated to how much participants were liked or respected. There was a main effect of other-impressions of agreeableness on respect (β = .06, p < .05). This suggests that those people who are agreeable are also respected when controlling for extraversion and moral impressions.
5 Study 4 discussion

The results of this study largely supported the optimal humility hypothesis. People whose self-impressions of moral character were humble compared to how they were seen by a close other were liked and respected more by that individual. These effects held when controlling for extraversion and agreeableness, suggesting that this relationship between an optimal amount of humility and social value is somewhat unique to the moral domain. Thus, people held less social value with a close other about their moral character when people overclaimed, were accurate, or extremely humble about their moral character. This was especially true when people enhanced their moral character relative to their close other’s impression.

In general, the evidence suggests an optimal level of humility. However, I have avoided presenting what the models have suggested is the optimal level because shared method variance between the informant impressions of morality and social value is likely inflating the degree of humility that is optimal. That is, part of what determines the optimal level of humility is the relative strength of the main effects of self-reports and informant reports in the polynomial regression equation.

While findings generally supported my optimal humility hypotheses, results for the four core virtues were nuanced. There was an optimal amount of humility regarding compassion and honesty when it comes to being liked and respected. But effects for loyalty were limited to being less liked and disagreement about fairness was not linked to social value. Although there was not much evidence for domain specific variance in Study 1, this pattern of results highlights the possibility that components of moral character might have different effects. Additionally, it suggests that researchers should measure more than one index of social value.

My findings add to the growing literature that self-enhancement in the eyes of others has social costs (Carlson & DesJardins, 2015; Colvin et al., 1995; Kurt & Paulhus, 2008; Paulhus, 1998). That said, I did not measure self-enhancement relative to an objective measure of moral character, and self-enhancement of moral character might be adaptive for the target in other ways. For example, people who hold positive illusions about their personality tend to experience greater well-being and report greater psychological adjustment (Church et al., 2006; Taylor & Brown, 1988; Taylor, Lerner, Sherman, Sage, & McDowell 2003), suggesting that while self-enhancement can have interpersonal costs, it can also have intrapersonal gains (Kurt & Paulhus, 2008; Paulhus, 1998). Thus, future work
might explore effects for self-enhancement relative to an objective criterion as well as if there are similar asymmetries in the costs and benefits of self-enhancement of moral character.

The current work also demonstrates the strengths of employing RSA rather than conventional approaches, such as difference scores or moderated regression. Indeed, difference scores often have problems with discriminant validity (Furr, 2011), and moderated regression approaches fail to capture the complexity of the various combinations of two predictors (Cafri et al., 2010; Meilich, 2006). In contrast, RSA retains information about both impressions when indexing discrepancy and reveals effects for all possible combinations of impressions. Indeed, past approaches would have been completely unable to test for optimal levels of humility which seems to best capture the effect of a shared understanding of morality and social value. Hopefully, future work exploring issues related to self-other agreement or discrepancies more broadly will also employ this approach.

While our results suggest that there is a robust link between self-other discrepancies about moral character and social value, we could not evaluate the causal direction of this link. In early acquaintanceship, individuals who tend to see themselves as others do are liked more over time than people who tend to self-enhance (Carlson & DesJardins, 2015; Human et al., 2012). Similarly, individuals who are more aware of how others perceive them are liked more over time than people who are less aware (Carlson, 2016). Thus, in this context, initial self-other disagreement might have led to less social value. Yet, it could be that losses in social value caused disagreement. For example, people might have bolstered their moral self-impression as a way of compensating for low social value. Future work employing a longitudinal or experimental design can shed more light on the direction of the effects we observed.

I focused on social value among close others, but future work might explore other interpersonal outcomes (e.g. influence, leadership, trustworthiness) and social contexts (e.g. coworkers, first impressions). In this study, targets were nominated and arguably liked by judges. In study 1 I found that the nature of moral impressions is consistent between liked and disliked targets which suggest the liked nature of targets should not be a major concern. However, future work examining outcomes in contexts where judges do not select their targets or in contexts where targets are not already liked (e.g. work groups) might yield stronger effects, given that judges who dislike targets perceive less positive qualities than do judges who like targets (Leising et al., 2010). Future research might also test the robustness of the extent to which these effects are unique to moral impressions by examining
the consequences of disagreement on similarly evaluative, but non-moral traits (e.g. competence, funny, intelligence).

Finally, my focus was on the judge’s experience of self-other discrepancies, but the target might also experience costs to such disagreements. For example, Tom might like Jane less when she sees his moral character differently than he sees his character or he might feel less authentic with Jane. Interestingly, some work suggests that the self experiences positive effects of enhancement while others experience negative effects (Kurt & Paulhus, 2008; Paulhus, 1998). Thus, it might be that the self must weigh the interpersonal costs against the intrapersonal benefits of enhancement.
Chapter 5
General Discussion

People quickly and naturally form moral impressions. However, researchers face many challenges when measuring these impressions because of the evaluative and internal nature of moral traits. The evaluative nature of moral traits suggests that part of moral impressions may arise because of more general evaluative biases of the perceiver. The internal nature of moral traits suggests that impartial judges may lack access to vital information about the morality of a person. Yet, despite these challenges, moral impressions are among the most consequential impressions people form. As such, moral impressions are a particularly interesting context to study self-knowledge and accuracy.

1 Extent of evaluative bias

To overcome the challenge of moral impressions being evaluative, I partitioned the variance in moral impressions into that which can be interpreted as a general evaluative bias, and that which is uniquely moral. In doing so, my results confirm the abundance of non-moral evaluative variance in moral impressions alongside moral impressions (ranging from 10-60%). There was an abundance of evaluative bias in all moral impressions. However, the amount of evaluative bias was particularly great in people’s impressions of others compared to people’s impressions of themselves.

As further evidence of the persistence of evaluative bias, it persists even in impressions from neutral judges who neither like nor dislike the person they are rating. Although evaluative bias might be expected from judges that like or dislike a person, these evaluative biases can’t be mitigated by using an impartial other. Further supporting the persistence of evaluative bias, the nature of these evaluative biases is the same regardless of the context of the impression. An inspection of the nature of the evaluative variance (i.e. the pattern of factor loadings) suggests that evaluative bias operates in similar ways between self- and other-impression, and in similar ways for judges at various levels of liking. This confirms the importance of modeling the sources of variance that contribute to moral impressions, but especially so when using other-impressions of a person.

2 Inter-judge consensus

Of the systematic variance in other’s perceptions, there was no evidence that a general moral impressions nor domain specific impressions are shared between two judges. Taken together, this
lack of effects suggests that a shared reputation about morality doesn’t form. Yet, past research suggests that a shared reputation does exist among perceivers (Helzer et al., 2014). In light of past work, my null findings are difficult to interpret. On one hand past work did not thoroughly partition evaluative biases from judges, thus it is possible that prior work overestimated the magnitude of consensus about morality. On the other hand, attrition limited the statistical validity of my consensus models. With little over 100 pairs of judges, it is very plausible that I did not have the statistical power to detect a real – but small – amount of consensus that exists.

3 Self-other agreement

From a measurement perspective, the existence of self-other agreement suggests that there is some validity to moral impressions. That is, systematic impressions of moral character are more than idiosyncratic ideas about how moral a person is. Instead, there is some truth to these impressions. Although shared impressions are not enough for demonstrating that these impressions reflect a valid ground truth, it is enough for demonstrating that a shared social reality forms about people’s morality.

Interpreting the practical significance of how much shared reality exists is tricky. By traditional standards a correlation of 0.23 would be considered a small to moderate effect (Cohen, 1988). But more a recent meta-analysis of individual difference correlations places a correlation of 0.23 well above the 50th percentile of observed correlations (Gignac & Szodorai, 2016). While these guidelines are helpful, neither are specific to the context of self-other agreement. We can further appreciate the meaning of this effect by placing it in the context of self-other agreement on Big Five traits and personality pathology. The observed agreement for moral impressions was below typical effects in the Big Five domain – including agreeableness (Connolly et al., 2007), instead similar to the size of agreement for personality disorders (Kaurin, Sauerberger, & Funder, 2018).

4 In lab and daily behaviors

Despite evidence of convergence between self- and other-impressions of morality, there was little evidence that self-impressions were related to moral behaviors. Self-impressions were not associated with in-lab cooperation behavior nor with daily reports from participants about their own behavior. There was some evidence that other-impressions are related to moral behavior. Specifically, other-
impressions were related to in-lab cooperation behaviors in the public goods game. However, there was no relationship between other-impressions and daily reports of moral behavior.

If my measures of behaviors are taken as a valid criterion of accuracy, then these results suggest that people are somewhat accurate about the morality of others. But that people are inaccurate about their own morality. This is mostly in line with SOKA model predictions that the evaluative nature of a moral impression prevents selves from holding accurate impressions and the somewhat internal nature of morality makes it difficult for other’s to form an accurate moral impression. And in general these results are in line with some recent work that suggest that impressions are not tethered to acoustical recordings of behaviors for either agreeableness (Beer & Vazire, 2017) or moral impressions (Bollich, 2016).

5 Is sharing reality adaptive?

Using polynomial regression and response surface analysis, I found that those participants that share a social reality did not experience the most social value. Instead, there was an optimal level of humility, such that those participants who somewhat underestimated their morality with respect to their peers were liked and respected the most. There seems to be a balancing act between the benefits of sharing a social reality and being somewhat humble with respect to peers. The result of this balance is that some, but not too much humility was associated with highest levels of social value.

Although the general pattern for moral impression suggested the optimal approach for participants was to be humble with respect to their peers, this was not the case for the domain of fairness. Instead, peer’s impression of fairness was positively related to social value, regardless of how people saw themselves. This suggests that there are no benefits of humility, nor costs to enhancement in the domain of fairness. Though some caution should be used when interpreting domain specific effects, as little systematic domain variance was detected in studies 1 & 2.

6 Implications for the study of moral psychology

Other-reports contained nearly double the amount of evaluative bias as self-reports. This suggests that in their raw form, moral impressions about other people may not be the most realistic impressions. This is particularly troubling given that much of the research in moral psychology uses participant’s ratings of strangers and hypothetical people in carefully crafted stories. If researchers
don’t disentangle global evaluative bias from moral impressions, then responses from participants about the morality of others is confounded by their impression of how positive they view the person in the story. In other words, researchers cannot be confident that ratings about people in carefully controlled vignettes reflect moral impressions or the evaluative lens that the participants have. This might be especially problematic when researchers experimentally manipulate the information raters have about targets, which may unknowingly change the evaluative lens that raters have but masquerade as real differences that are specific to morality.

Aside from the domain compassion, less than 10 percent of the variance in moral items was systematic and specific to a given moral domain. Instead, much of the systematic variance can be explained by a general moral impression, or general evaluative tendencies. As such, efforts to study specific moral domains through surveys or experiments must carefully design studies in order to extract what little signal remains. Large samples will be required to have the statistical power to detect signal from these relatively small sources of variance. Further clear manipulation checks that also probe non-moral evaluative traits will be necessary to confirm that a particular domain of morality and not morality in general or other evaluative traits have similarly been manipulated by experimental designs and thus confound the researchers conclusions.

7 Limitations and future directions

When testing the extent to which moral impressions are grounded to moral actions, I examined the link between moral impressions and the action of cooperating in a public goods game. This was a very conservative test of the link between moral impressions and actions because impressions of the whole person are rarely tied to one specific action. People’s morality is likely to fluctuate greatly from situation to situation, yet still manifests itself in typical behaviors. There are at least two reasons why my observed correlations might underestimate the actual relationship between impressions and patterns of behavior. First, there is generally more intra personal variance than between person variance (Fleeson, 2001; Fleeson, 2004). That is, people stray more from typical behavior from moment to moment, than people stray from the typical person’s behavior. The presence of all this intrapersonal variance makes people behave quite different from moment to moment and from situation to situation. Thus, any one observation of behavior is not a stable estimate of the typical behavior of a person. Second, the laboratory setting can be a particularly strong situation which prevents between-person variability from expressing itself. When we think
about what drives any behavior, it is undoubtedly the consequence of both the person and the situation. Some situations exert very strong pressures which constrain the influence of personality on behavior. The laboratory setting is likely one of these settings (Diener, Larsen, & Emmons, 1984). With these limiting factors in mind, future work seeking to establish the accuracy of moral impressions would benefit from repeated measures of moral behaviors across different settings. Ideally, this work would include multiple assessments of both laboratory and naturalistic settings.

All the impressions collected in my work were impressions of real people. This is largely because people naturally form impressions of the whole person. As such, my work has strong external validity. However, a consequence of exclusively measuring real and whole persons is that my work does little to explain the judgements of specific actions. In moral psychology researchers measure impressions of the morality of people, but also of specific actions. Indeed, whole areas of psychology and philosophy are dedicated to understanding what are the features of an action that make it moral or immoral. Or what makes the agent of an action morally blameworthy for the outcome of a specific action. In my work I did not measure any impressions of actions, thus it remains to be tested if my measure of moral impressions can or should be used to measure the morality of specific actions. It is entirely plausible that people’s judgements of specific actions do not follow the same structure as impressions of whole persons. Future work is necessary to bridge the gap in our understanding of the ways in which judgments of people and actions are similar or dissimilar. For example, are judgments of a person’s morality the product of judgements of their actions, are judgements of a person’s action impacted by preexisting impressions of the actor’s morality, or do moral impressions of actions and people reciprocally influence each other?

8 Conclusion

In conclusion, when people provide their impressions of their own or other’s morality, that impressions contains a mix of their moral impression and broad evaluative biases. These results suggest that researchers should be extremely cautious about interpreting raw moral impressions, because moral impressions are infused with evaluative variance that is not unique to morality or its domains. After controlling for these evaluative biases, the uniquely moral impressions that people form seem to be somewhat tethered to a social reality. Those people who see themselves as moral tend to be seen as moral by others. And the magnitude of agreement on a moral impression is similar to other traits which are difficult to judge. Other-impressions were somewhat tethered to in-lab
behavior, but self-impressions were unrelated to moral behavior. Finally, sharing a social reality wasn’t optimal for achieving liking and respect among peers. People’s social value seems to be optimized when they hold a somewhat humble impression of their own morality compared to those around them.
References


https://openscholarship.wustl.edu/art_sci_etds/832


Appendices A: List of Scales Included in Study 1

Included scales:

- Personality Judgments
- Cognitive Reflexive Task
- Moral Foundations
- The Personality Inventory for DSM-5
- Philadelphia Mindfulness Scale
- Inventory of Interpersonal Problems
- Demographics

Attention Check Items (Correct answers bolded)

- How many U. S. states are there? **50**
- Christmas Falls in which month of the year? **12th**
- Enter four to indicate that you are paying attention to the questions. **4**
- How many moons does earth have? **1**
- Select the highest of the numbers below. **5**
- To what degree are you paying attention to the questions? **4,5**
- I am randomly responding. **False**
- Who is the current president of the US? **Barack Obama**

What season begins after summer? **Fall/Autumn**
Appendices B: List of Scales Included in Study 2

Included scales:

- Personality Judgments
- Big Five Inventory
- Personality Inventory for DSM-5 Brief Form
- Short Moral and Personality Items
- Evaluative Items
- Single-item Self-Esteem Scale
- SIAS-6
- NPI-13
- Self-Compassion Scale
- Revised Self-Monitoring Scale
- MAAS
- Philadelphia Mindfulness Scale
- Five Facet Mindfulness Questionnaire
- MAIA
- Balanced Inventory of Desirable Responding – Short Form
- Social Desirability Scale – Short Form
- Ruminative Response Scale
- Mindfulness Knowledge and Experience Questions
Appendices C: Items in Socially Desirable Responding Scales

**Balanced Inventory of Desirable Responding (BIDR)**
1. I have not always been honest with myself.
2. I always know why I like things.
3. It's hard for me to shut off a disturbing thought.
4. I never regret my decisions.
5. I sometimes lose out on things because I can't make up my mind soon enough.
6. I am a completely rational person.
7. I am very confident of my judgments.
8. I have sometimes doubted my ability as a lover.
9. I sometimes tell lies if I have to.
10. I never cover up my mistakes.
11. There have been occasions where I have taken advantage of someone.
12. I sometimes try to get even rather than forgive and forget.
13. I have said something bad about a friend behind his or her back.
14. When I hear people talking privately, I avoid listening.
15. I never take things that don't belong to me.
16. I don't gossip about other people's business.

**Social Desirability Scale (SDS)**
Items:
1. It is sometimes hard for me to go on with my work if I am not encouraged.
2. I sometimes feel resentful when I don't get my own way.
3. On a few occasions, I have given up doing something because I thought too little of my ability.
4. There have been times when I felt like rebelling against people in authority even though I knew they were right.
5. No matter who I'm talking to, I'm always a good listener.
6. There have been occasions when I took advantage of someone.
7. I'm always willing to admit it when I make a mistake
8. I sometimes try to get even, rather than forgive and forget.
9. I am always courteous, even to people who are disagreeable.
10. I have never been annoyed when people expressed ideas very different from my own.
11. There have been times when I was quite jealous of the good fortune of others.
12. I am sometimes irritated by people who ask favours of me.
13. I have never deliberately said something that hurt someone's feelings.
Appendices D: List of Scales and Tasks in Study 3

List of Scales:

- Motivation Action Character Scale (MACS)
- Belief In a Just World Scale
- Moral Foundations Questionnaire (MFQ)
- Circumplex Scales of Interpersonal Values
- Philadelphia Mindfulness Scale
- Pathological Narcissism Inventory
- Wonderlic Scale
- The Personality Inventory for DSM-5 (PID5)
- Self-Control Scale
- Tangney Self Control Scale
- Cognitive Reflection Test (CRT)
- Brick Creativity Task
- Flanker Task
- Photo Taken
- Public Goods Game
- Man on Moon Task
- Video Rating Task
- Demographics