**In Defense of Ordinary Moral Character Judgment**

Evan Westra

**Abstract:** Moral character judgments pervade our everyday social interactions. But are these judgments epistemically reliable? In this paper, I discuss a challenge to the reliability of ordinary virtue and vice attribution that emerges from Christian Miller’s Mixed Traits theory of moral character, which entails that the majority of our ordinary moral character judgments are false. In response to this challenge, I argue that a key prediction of this theory is not borne out by the available evidence; this evidence further suggests that our moral character judgments do converge upon real psychological properties of individuals. I go on to argue that this is because the evidence for the Mixed Traits Theory does not capture the kind of compassionate behaviors that ordinary folk really care about. Ultimately, I suggest that our ordinary standards for virtue and vice have a restricted social scope, which reflects the parochial nature of our characterological moral psychology.

**1. Introduction**

Since G.E.M. Anscombe’s seminal paper, “Modern Moral Philosophy,” the study of moral character has enjoyed a renaissance in philosophy (Anscombe, 1958). In the past two decades, moral character has also attracted a great deal of additional interest from empirically minded philosophers and moral psychologists. Two different kinds of research project have emerged from this increased attention. The first kind of project has been concerned with the nature of the moral character traits that we actually possess, and the extent to which they reflect the traditional Aristotelian virtues and vices. Most memorably, situationists about moral character argued that most people do not possess stable, cross-situationally consistent, “global” character traits at all, and thus the traditional Aristotelian conception of moral character is empirically false (Alfano, 2013; Doris, 2002; Harman, 1999). This provoked a wave of replies that challenged the empirical and philosophical bases of the situationists’ argument (Jayawickreme, Meindl, Helzer, Furr, & Fleeson, 2014; Kamtekar, 2004; Paris, 2016; Sabini & Silver, 2005), as well as a number of detailed, empirically grounded positive accounts of moral character and global character traits (C. B. Miller, 2013; Russell, 2009; Snow, 2010). At this time, philosophers and personality psychologists seem generally optimistic about developing a robust science of moral character (Fleeson, Furr, Jayawickreme, Meindl, & Helzer, 2014; C. B. Miller, 2017).

The second research project has been concerned with the role that *concepts* of moral character and moral character traits play in our everyday moral judgments and social evaluations. Strikingly, our judgments about moral character often diverge from our judgments about the moral permissibility of certain actions. For example, while most people might judge hitting a coworker in a fit of anger to be more morally impermissible than calling a coworker a racial slur under one’s breath, the latter was viewed as much more indicative of a bad moral character (Uhlmann, Zhu, & Diermeier, 2014). Indeed, beliefs about moral character appear to play a dominant role in our overall impressions of other people (Goodwin, Piazza, & Rozin, 2014), as well as our conceptions of personal identity (Strohminger & Nichols, 2014), and even in the way we interpret people’s intentions (Sripada & Konrath, 2011). These and other results have led a number of authors to suggest that folk moral judgment embodies a kind of “intuitive virtue ethics”, reflecting an evolutionary adaptation for selecting reliable cooperative partners (Landy & Uhlmann, 2018; Martin & Cushman, 2015; Uhlmann, Pizarro, & Diermeier, 2015).

The focus of this paper concerns a set of questions at the intersection of these two research projects about moral character: to what extent do our intuitive moral character judgments reflect the character traits that people actually have? That is, are ordinary moral character judgments – and especially attributions of virtue and vice – at all accurate or reliable?

For highly skeptical views like situationism, which hold that most people possess highly local, situation-specific character traits, the answer to these questions are inevitably pessimistic: if the vast majority of people do not possess global traits, there is little hope for accurate or reliable global trait attributions (Alfano, 2011; Doris, 2002; Harman, 1999). But even for views that accept the widespread existence of global character traits, it is not obvious that moral character judgment should be viewed as trustworthy or reliable. One argument along these lines comes from Christian Miller’s Mixed Traits theory of moral character (henceforth, MTT) (C. B. Miller, 2013, 2014). The central claim of this theory is that although the traditional virtues and vices exist, they are statistically rare in the general population. Rather, the vast majority of individuals possess morally heterogeneous *Mixed Traits*. Mixed Traits are global character traits grounded in stable and consistent psychological dispositions that nevertheless cause us to act in ways that appear virtuous in certain situations and vicious in others. Crucially, these Mixed Traits do not correspond to any of our folk moral character trait concepts, such as the traditional virtues and vices. Thus, there is a fundamental mismatch between the moral trait categories that we apply in our everyday moral character judgments and the traits people actually possess. This means that the vast majority of our ordinary virtue and vice attributions are false – not because global character traits and the virtues and vices do not exist, but because of a conceptual deficit in our moral character judgment.

In this paper, I aim to defend the adequacy of our ordinary moral character trait concepts and the reliability of ordinary moral character judgment. I begin by presenting a set of data that is *prima facie* problematic for MTT: while the MTT argues that we mostly err in our ordinary moral character attributions, there is a substantial body of evidence showing that these judgments consistently agree with one another, and reliably predict real life outcomes. This suggests that our moral character judgments do reliably track real psychological properties of individuals, which runs contrary to what the MTT seems to predict. To explain the source of this failed prediction, I argue that the evidence upon which the MTT is based does not reflect the standards for virtue and vice that most ordinary folk would adopt in their everyday moral character judgments. The resulting account avoids the difficulties that come along with an error theory about ordinary moral character judgments, and also sheds some light on the social functions that these judgments play in our everyday lives.

**2. The MTT and the CET**

To understand why Miller argues that most of us do not possess the traditional virtues and vices, we must begin by outlining his framework for understanding the nature of character traits in general. Like many other virtue theorists and personality psychologists, Miller thinks that character traits are dispositions to form certain kinds of mental states in certain types of trait-relevant situations; all things being equal, these mental states should also lead to trait relevant behaviors (C. B. Miller, 2013, p. 6, 2014, p. 6). For example, the trait of extraversion might involve the disposition to prefer larger social groups, to believe that one is skilled in such situations, and to desire to be the center of attention; all things being equal, these dispositions ought to lead a person to behave in a generally extraverted manner.[[1]](#footnote-1) Miller holds that character traits are constituted by these mentalistic dispositions. There is nothing at all to being a character trait over and above being disposed to have certain kinds of mental states.

*Moral* character traits are a subset of character traits that affect our moral evaluations of a person. For example, how *honest* a person is affects how we morally evaluate that individual, but how *witty* a person is does not; hence, honesty is a moral character trait, but wittiness is not. There are many classes of moral character trait, associated with different domains of morally relevant action (e.g. harming, helping, lying, stealing, etc.). Within each of these classes, we can distinguish between three different types of trait: virtues, vices, and Mixed Traits. For example, there are three types of trait associated with the domain of harming: harming-relevant virtues (e.g. non-malevolence), harming-relevant vices (e.g. cruelty), and what Miller dubs “Mixed Aggression Traits” (C. B. Miller, 2014, p. 45).

Virtues and vices are partially defined in terms of our commonsense platitudes about the behaviors that virtuous or vicious people typically engage in, and the motivating reasons that typically underlie those behaviors. For example, Miller writes of the virtue of compassion that,

A person who is compassionate, when acting in character, will typically attempt to help when, at the very least, the need for helping is obvious and the effort involved in helping is very minimal. (C. B. Miller, 2013, p. 190)[[2]](#footnote-2)

These platitudes are framed as necessary conditions or *minimal thresholds* for virtue and vice. They do not describe the behaviors or attitudes of moral (or immoral) exemplars, individuals who possess the virtues (or vices) to a very high degree. They merely aim to capture the most basic conditions for true attributions of virtue and vice.

The bulk of Miller (2013) is devoted to a thorough survey of the empirical literature[[3]](#footnote-3) on factors that affect motivation to engage in compassionate, helping behavior, which serves as the primary case study for the MTT. One broad theme of this survey, which holds across dozens of studies, is that most people do not engage in even moderately demanding helping behavior. A second theme is that people are often more likely to help in the presence of motivating factors that are less than virtuous. These include the desire to alleviate feelings of guilt or embarrassment, the desire to continue one’s good mood or alleviate one’s bad mood, and the desire to appear morally good to one’s self and others. Sometimes we do help for genuinely altruistic motives, but this tends to depend upon the presence of positive moral enhancers, such as empathy. But the tendency to empathize is both fragile and fickle, and thus does not amount to a stable virtuous disposition. In short, most people fail to reliably engage in helping behavior; when we do help, it is often for bad reasons.

Thus, Miller concludes, most of us do not possess the traditional virtue of compassion. Instead, he suggests that most of us possess what he calls *Mixed Helping Traits*, which are constituted by a morally heterogeneous mix of psychological tendencies, such as the disposition to help when one feels guilty or when it is in one’s self interest, or to not help when one fears embarrassment, or to help when experiencing fleeting bouts of empathy. Collectively, these dispositions lead to behavior that can appear morally inconsistent to an observer. But they actually represent quite stable features of a person’s psychology, and vary between individuals.

Miller goes on to argue that this pattern of empirical results generalizes across other domains of moral behavior, such as aggression, lying, and cheating. In none of these domains do we find people behaving in a manner consistent with the relevant virtues *or* vices. Instead, we find the same pattern of moral heterogeneity. Across the board, most people act in ways that are consistent with the widespread possession of Mixed Traits.

Crucially, Miller claims that Mixed Traits do not correspond to any of our ordinary moral trait terms or concepts, such as the traditional virtues or vices (C. B. Miller, 2013, p. 156). There is thus a basic mismatch between the nature and distribution of moral character traits in the world on the one hand, and our trait-related conceptual resources on the other. Because of this concept-trait mismatch, it is actually impossible for us to form true beliefs about Mixed Traits as such with the conceptual resources we currently possess (though presumably we could accurately describe the individual psychological dispositions that constitute these traits). This basic conceptual deficit is exacerbated by the fact that we are prone to a psychological bias known as the “fundamental attribution error,” which causes us to attribute traditional virtues and vices to people on the basis of single observations of morally relevant behavior (C. B. Miller, 2014, p. 167).[[4]](#footnote-4) Thus, if we see a person engage in one courageous act, we tend to infer that she is courageous; if we see her engage in a cruel act, we tend to infer that she is cruel (when in both cases her actions most likely reflect Mixed Traits). This means that we both lack the conceptual capacity to form true beliefs about Mixed Traits, *and* we are also psychologically disposed to form false beliefs about them. Miller calls this part of his view the Character Error Theory, or CET.

The CET leads to a practical dilemma about our ordinary trait-attribution practices: should we advocate for the elimination of virtue- and vice-attribution from our daily lives, and replace it with a new vocabulary about Mixed Traits? Or should we persist in falsely attributing virtue and vice? This dilemma is complicated by the fact that virtue-attributions can be an effective way of encouraging positive moral behavior (Alfano, 2013, pp. 88–108). Thus, while eliminativism about virtue and vice talk would make sense from an epistemic perspective, it might also have negative moral consequences. Retaining virtue and vice talk as is would lead to the opposite trade-off. Thus, neither option is particularly appealing.

**3. Agreement about moral character**

Note that the MTT does not deny that virtues and vices *exist*: it merely states that such traits are statistically rare. It also allows that we are in fact correct when we describe moral exemplars like Ghandi and Martin Luther King Jr. as virtuous, or when we describe immoral exemplars like Hitler and Stalin as vicious. To this extent, the MTT acknowledges that most of us have a basic capacity to successfully pick out real instances of virtue and vice at least some of the time, when provided with sufficiently strong positive evidence. Thus, CET is unlike the sort of error theory that (say) an atheist might invoke to explain a theist’s belief in an all-powerful, all-knowing, benevolent God. According to the atheist, the theist who talks about God always fails to refer to anything. According to the CET, we can sometimes talk about virtue and vice and get it right.

The real “error” in the CET lies in the fact that very often, we *miscategorize* people who possess Mixed Traits as virtuous or vicious. When I call my friend who regularly gives to charity “generous”, the aspect of her moral character that I am in fact responding to is something like a Mixed Giving Trait. This trait consists in a set of psychological dispositions that lead my friend to give for morally laudable reasons on some occasions, to give for morally suspect reasons on other occasions, and sometimes to fail to give when she in fact had good reason to do so. Although some of the dispositions that constitute my friend’s Mixed Giving Trait are consistent with the virtue of generosity (hence my error), on the whole, her trait does not meet the minimum normative standards for virtue. Thus, when I call my friend “generous”, I am in fact making a mistake.

One might expect that such errors would lead to various kinds of practical confusion and disagreement about which moral character traits specific people possess. To see why, consider the following scenario:

**Outbreak**: Imagine a community in which there has been an outbreak of a strange new virus. The doctors in this community only know of two different types of virus: V1, with symptoms A and B, and V2, with symptoms C and D. Unbeknownst to them, V1 and V2 are in fact quite rare within the population. However, the new virus V3 sometimes leads to symptom B, and sometimes to symptom C. When doctors see patients experiencing symptom B, they will diagnose their illness as V1; when they see a patient suffering from C, they will diagnose their illness as V2. Because the same patients sometimes display B and C, if they go to different doctors, they would be likely to receive multiple contradictory diagnoses.

The situation in Outbreak is analogous to what we face if the MTT is correct. For instance, if most ordinary folk possess Mixed Aggression Traits, but we only possess concepts for the full-blown virtue of kindness and the full-blown virtue of cruelty, then every time we encounter a person with a Mixed Aggression Trait, we will inevitably “misdiagnose” it. The same person might appear kind to one observer and cruel to another, or kind to an observer on one occasion, and cruel to that same observer on a different one. Because of the inadequacy of our trait concepts, we should expect to find widespread disagreement (and confusion) about who is virtuous and vicious. Like the doctors constantly misdiagnosing their patients, under the Mixed Traits hypothesis, we simply lack the appropriate concepts to carve up the characterological world at its joints.

This prediction of the MTT creates a testable hypothesis: if we systematically miscategorize people with Mixed Traits as virtuous or vicious, we should expect widespread disagreement about which moral character traits people have. Fortunately, measuring levels of agreement about trait attributions has been studied by psychologists for decades (Connelly & Ones, 2010; Funder, 1995; Kenny, 1991). The basic method for studying “accuracy in personality judgment” involves giving one or more informants or judges personality tests, which they must fill out about a particular target individual (sometimes someone they have met in a laboratory context, and sometimes a person they know in their everyday lives). Researchers can then analyze the degree to which the personality trait ratings of different judges correlate with one another (“inter-informant consensus”), and how well self-reported personality ratings correlate with those of third-party observers (“self-other agreement”). Many factors have been found to affect these correlations, such as the relationship between the judge and the target and the kind of trait being rated (Biesanz, West, & Millevoi, 2007; Connelly & Ones, 2010; Funder, 1995; Vazire et al., 2010). But in general, agreement about other people’s personality traits is a very robust phenomenon.[[5]](#footnote-5)

Some of this research bears upon the question of whether people disagree about moral character. While personality questionnaires using the Five Factor model of personality do not specifically target moral character, two of the Five Factors, Agreeableness and Conscientiousness, are made up of facets that plausibly represent moral traits, such as “Modesty,” “Altruism,” and “Dutifulness.” Meta-analyses of studies measuring inter-informant consensus on Agreeableness and Conscientiousness reveal moderate positive correlations (*rs* = .4 and .44), and strong correlations for self-other agreement (*rs* = .71 and .82), suggesting there is at least modest but consistent agreement about these traits (Connelly & Ones, 2010, pp. 1102–1106).

However, a downside of using personality tests based on the Five Factor Model to measure moral character (either first-personally or third-personally) is that these tests were not designed for that purpose; indeed, as Miller points out, many of these questionnaires’ putatively moral facets such as “altruism” seem to measure decidedly non-moral factors, such as how people think of the target, and the target’s manners and politeness (e.g. “Most people I know like me” and “I try to be courteous to everyone I meet”) (Costa & McCrae, 1992; C. B. Miller, 2014, p. 143). This is not to say that answers to these questions do not relate to moral character in some way: several of the questionnaire items *do* seem to be tracking morally relevant aspects of personality (e.g. “I go out of my way to help others if I can”) (Costa & McCrae, 1992; C. B. Miller, 2014, p. 143). But at most, they should be treated as an indirect measurement of consensus about moral character.

These concerns about Five Factor Model do not apply to the HEXACO model of personality, however (Ashton & Lee, 2007; Lee & Ashton, 2008). HEXACO adds a sixth “Honest-Humility” factor to the Big Five personality dimensions, which “represents the tendency to be fair and genuine in dealing with others, in the sense of cooperating with others even when one might exploit them without suffering retaliation” (Ashton & Lee, 2007, p. 156), and is reflected in adjectives such as *sincere, honest, faithful, modest, sly, greedy,* and *hypocritical* (Ashton & Lee, 2007, 2009; Ashton, Lee, & Goldberg, 2004; Lee & Ashton, 2008). Questionnaire items for this personality factor are much more distinctively moral in their content. For example, items for the Honest-Humility factor in the HEXACO-60 include prompts like “I wouldn’t pretend to like someone just to get that person to do favors for me,” and “I’d be tempted to use counterfeit money, if I were sure I could get away with it [reverse-scored]” (Ashton & Lee, 2009). Self-reported scores on the Honesty-Humility dimension have also been shown to predict unethical behaviors and decision making in many contexts (Dunlop, Morrison, Koenig, & Silcox, 2012; Hilbig & Zettler, 2009; Lee, Ashton, & De Vries, 2005). It is thus quite plausible that questionnaires based on the HEXACO model reflect morally relevant aspects of a person’s character.[[6]](#footnote-6)

Most studies measuring agreement on the HEXACO dimensions have focused on self-other agreement. In populations of well-acquainted college students, self- and other-ratings of Honesty-Humility have been found to be highly correlated (*r* = 0.52-0.54) (Ashton & Lee, 2010; Lee et al., 2009; Lee & Ashton, 2017). In two studies with more diverse populations, Taya Cohen and colleagues found similarly high levels of self-other agreement for this personality factor (*r* = 0.43-0.5), as well as for ratings of “Guilt-Proneness,” another personality measure that correlates with Honesty-Humility and predicts unethical behavior (Cohen, Panter, Turan, Morse, & Kim, 2013; Cohen, Wolf, Panter, & Insko, 2011). These correlations remain comparable even after controlling for variance explained by judges’ own self-reports (Cohen et al., 2013; Lee et al., 2009), and are not likely to be the product of judges responding in a socially desirable way (Roth & Altmann, 2019). Correlations between self- and other-ratings of Honesty-Humility also tend to be higher in dyads that are better acquainted with one another (Lee & Ashton, 2017). These results indicate that self-other agreement on Honesty-Humility is quite high, and that it increases as judges to know their targets more intimately.

The aforementioned studies all measured agreement about moral character by comparing absolute ratings on particular trait dimensions. Another way to assess agreement about moral character (and personality more generally) to measure agreement about moral trait *profiles* (Furr, 2010). A profile-based approach to captures whether or not different judges agree about a particular individual’s relative moral strengths and weaknesses, even if they do not agree in their absolute ratings for specific traits. For example, one person might think that Gina is very temperate, while another might think that she is only moderately temperate. However, they might both agree that she is more compassionate than she is temperate, and more loyal than she is compassionate. Helzer and colleagues did just this using an original “Moral Character Questionnaire,” which included items meant to elicit judgments about targets’ virtues: fairness, honesty, compassion, temperance, “moral concern”, as well as measures for “general morality” and “moral character” (Helzer et al., 2014). Although they found relatively lower (though still significant) levels of agreement on individual trait-ratings than previous studies (*r* = .15-.42) (Helzer et al., 2014, p. 1703), Helzer and colleagues also found strong profile-level agreement for both self-other agreement and inter-informant consensus (*r* = .73 and .66) (Helzer et al., 2014, p. 1705).[[7]](#footnote-7) Again, this suggests that agreement about moral character is fairly robust.

Notably, these data do not show that there is perfect consensus about moral character: there is still a fair amount disagreement in how people perceive each other’s moral traits. But some disagreement is inevitable given the challenges that are inherent to personality judgment generally. First, personality (and the mentalistic dispositions that undergird it) is not something that can be directly observed. It must be inferred from observations of a person’s behavior. But how and whether a particular trait manifests itself in a person’s behavior very much depends on the situation a person is in. A person may have brave dispositions, for instance, but these may not manifest themselves in the behavior observed by a given judge if the judge does not witness that person in a bravery-relevant situation (e.g. standing up to a bully). Moreover, the features of the situation that are salient to a third-party informant (or “nominal situation”) may not align with the features of the situation that are most salient to the target (the “psychological situation”) (Mischel, 2004; Mischel & Shoda, 1995). In short, inferring character from behavior is a hard and noisy process, and there are a number of noise factors that can be expected to drive down agreement about character.

The MTT predicts that in addition to all these noise factors, we also employ an inadequate set of moral character concepts to categorize behavior; in other words, the Outbreak-like dilemma is layered *on top of* all the challenges inherent to the character judgment process. Given this prediction, it would be extremely surprising to see the levels of agreement that we in fact observe. This suggests that the criteria for attributing certain virtues and vices are comparable across different judges; when applying these standards, different judges tend to converge upon similar judgments about particular individuals, even when operating with distinct observational bases. Unlike in the case of Outbreak, where the doctors’ inadequate V1 and V2 concepts prevented them from reliably categorizing the real illness people were suffering from, our ordinary virtue and vice concepts *do* permit us to reliably categorize people. There is agreement about which individuals belong in the extensions of different virtue and vice terms.

This interpersonal consistency also provides compelling evidence that our virtue and vice attributions are converging upon real psychological properties of individuals. In the philosophy of science, convergence upon similar measurements across independent observations is commonly taken to reflect the reality of the entity being measured – a phenomenon William Whewell called a “consilience of inductions” (Helgeson, 2013; Whewell, 1989). Consensus about moral character judgments across independent judges seems to be an instance of this phenomenon: different people, presented with different behavioral evidence, still reach the same conclusion about a given target’s underlying traits. In short, our ordinary virtue and vice attributions are onto *something* (Westra, 2019).

One way for the Mixed Trait theorist to respond to this evidence would be to suggest that agreement about moral character does not reflect any underlying characterological realities in the targets, but rather shared stereotypical beliefs in the judges themselves. For instance, if two judges both hold the sexist stereotypes that women tend to be kind, and men tend to be brave, then they will be more likely to converge upon kindness judgments about women and bravery judgments about men. Given the pervasiveness of these kinds of stereotype (Fiske, Cuddy, & Glick, 2002), it may be that what agreement there is about ordinary folks’ moral character traits is entirely attributable to them.

Two considerations speak against this objection. First, third-party ratings of “moral” personality trait dimensions such as Conscientiousness and Agreeableness tend to be reliable predictors of important life outcomes, above and beyond self-reports (Luan et al., 2018; Vazire et al., 2010; Vazire & Carlson, 2011). Third-party ratings on the “Honesty-Humility” factor of the HEXACO personality measure also reliably predict targets’ fair distributions in Dictator Games (Thielmann, Zimmermann, Leising, & Hilbig, 2017). This would not be the case if these judgments were not tracking real psychological properties. Second, the degree of agreement about such traits tends to increase as a function of how well judges are acquainted with the target: the better two people know a target, the more their judgments about the target’s personality traits are likely to agree (Biesanz et al., 2007; Connelly & Ones, 2010; Funder, 1995). If judges’ character ratings were driven solely by widely shared stereotypes, then we should not expect to find that additional experience with the target should make a difference to how judges perceive them. If, on the other hand, judges are reliably tracking aspects of targets’ actual moral dispositions, then this is exactly the pattern of results that we should expect to find. This is not to say that stereotypes or implicit biases never distort our judgments about moral character, but rather that these distortions cannot fully explain the patterns of agreement about moral character that we do observe (Westra, 2019).

To summarize: if the CET were correct, our commonsense attributions of virtue and vice should only converge when we make character judgments about individuals at the extreme ends of the moral character spectrum. Towards the middle of that spectrum, our judgments of character should be all over the map. Instead, when we deploy our folk concepts of virtue and vice, we tend to agree with each other. What’s more, that agreement seems to be based on the character traits that people actually have. Where the CET predicts disagreement, we find agreement. Where the CET predicts inconsistent categorization, we find consistency. This is not what widespread error is supposed to look like.

**4. Compassion in context**

The above argument leaves us with a puzzle: on the one hand, the MTT argues that most ordinary people do not live up to certain minimal, commonsense thresholds for virtue and vice, which suggests that most of our ordinary moral character judgments must be erroneous and confused. On the other hand, our actual virtue and vice attributions don’t seem to bear the empirical signatures of error or confusion.

One possible strategy for resolving this puzzle would be to argue that the MTT’s minimal thresholds for virtue and vice are too high, and do not actually reflect commonsense standards for virtue and vice. In reply, a proponent of the MTT could reasonably argue that its minimal thresholds already capture the absolute lowest standards for virtue and vice consistent with ordinary usage. To see why this response would be reasonable, consider again one of Miller’s behavioral thresholds for compassion:

A person who is compassionate, when acting in character, will typically attempt to help when, at the very least, the need for helping is obvious and the effort involved in helping is very minimal. (C. B. Miller, 2013, p. 190)

It is hard to see how this standard could be weakened. Perhaps one could change the word “typically” for “occasionally” – but it is difficult to imagine anyone’s concept of compassion answering to such a low bar. It is equally difficult to imagine requiring anything less than attempting to help when the need for it is “obvious and the effort involved… is very minimal.” If Miller is right, and hundreds of empirical studies show that the majority of people do not meet these standards, then it is not obvious how such individuals could count as compassionate.

Thus, the source of the puzzle is not that the MTT’s standards for virtue are too high, or that it does not capture our commonsense understanding of compassion; we can grant that it does so fairly well. It is rather that the empirical results in question *do not in fact show* *that most people fail to meet these standards*. This is because our ordinary commonsense standards for compassion are based on our observations of how compassionate and uncompassionate people *typically* think and act; indeed, the qualifier “typically” is built right into the MTT’s minimal standards. To meet our ordinary criteria for compassion, then, a person need not act compassionately across all compassion-relevant situations. In particular, compassion-inconsistent actions in situations that are neither typical nor relevant would not show that a person lacks compassion according to commonsense standards. What will really matter for ordinary moral character judgments is whether a person is compassionate in typical compassion-relevant situations – the kind of situations that we care about in practice. In order to test whether most people meet these standards, the studies in question must therefore reflect such typical compassion-relevant situations. And there is good reason to believe that this is not the case.

To see why, consider how the studies in question are usually designed: a naïve participant is placed in a context in which a confederate of the experimenter suddenly requires some form of assistance, or in which they can make a small donation to a charity; the dependent measure is always related to whether the participant attempts to provide this assistance or makes the donation (Batson et al., 1989; Cann & Blackwelder, 1984; Freeman, Aquino, & McFerran, 2009; Latané & Rodin, 1969; Regan, Williams, & Sparling, 1972). While other aspects of the situation may vary a great deal, a common feature remains the *lack of acquaintance* between the participant and the individual in need of help. In effect, what these studies test is thus not helpfulness in general, but rather helpfulness towards strangers. Thus, the MTT relies upon measures of compassion in *contexts of zero-acquaintance* to establish whether people typically meet the standards for compassion.

Now, consider how levels of acquaintance impact helping behavior. In the context of existing interpersonal relationships, most of us probably engage in helping behavior quite often. We regularly drive family members to the airport, help friends move, comfort sad coworkers, do small favors for our neighbors, and so on. And, of course, some people will help more in these kinds of contexts than others. Not coincidentally, these are also the kinds of context in which we observe other people manifest varying levels of compassion. This is the kind of evidence we are likely to use when judging whether a particular person we know is compassionate. Such a judgment plausibly involves considering how we have observed this person act in compassion-relevant contexts, and perhaps a tacit comparison with how we have observed others act in similar circumstances. Based on this comparison, we might estimate whether a person is more or less compassionate than average, and whether they meet the standard for compassion. In other words, ordinary compassion judgments are based on observations of compassion-relevant behaviors that happen *in the context of existing interpersonal relationships*.

This is no mere conjecture: many aspects of human social cognition are calibrated relative to the number of close personal relationships that we regularly maintain, from prefrontal cortical volume, to measures of social competence, to empathic responses to others’ pain (Dunbar, 2008; Meyer et al., 2013; Powell, Lewis, Roberts, García-Fiñana, & Dunbar, 2012; Stiller & Dunbar, 2007). Within these networks (which, in humans, average around 150 individuals), we are highly partial in how we expend our socio-cognitive resources: most people devote roughly 40% of their available social time to their 5 most intimate friends and relatives (or “support clique”), and 20% of their time to the next 10 individuals with whom they are closest (or “sympathy group”); in other words, we spend 60% of our available social resources on just 15 individuals (Sutcliffe, Dunbar, Binder, & Arrow, 2012). Not coincidentally, it is within these inner circles of the social network that we lend one another more substantial forms of emotional and instrumental support (Dunbar, 2014).

It is thus quite plausible that these kinds of relational context are the ones that are most relevant to us in our everyday evaluations of moral character. This fits nicely with theories about the adaptive function of moral character judgment – namely, to select reliable cooperating partners, and thereby protect one’s self from exploitation by free-riders (Baumard, André, & Sperber, 2013; Landy & Uhlmann, 2018; Martin & Cushman, 2015). According to these “partner choice” models of the evolution of cooperation, individuals can limit their exposure to exploitation by choosing to opt in or out of cooperative endeavors with potential partners on the basis of whether that individual is likely to be a reliable cooperator. Reliable assessments of moral character enable us to do this by providing a means to infer whether an agent is disposed to act in a prosocial or antisocial manner. Ultimately, this creates a selective pressure against freeriding and towards mutualistic cooperation within the broader population. Given that most high-stakes cooperation occurs in the context of close interpersonal relationships, there is an obvious selective pressure for individuals to base their character judgments on behaviors that occur in those social contexts. That is, from a partner-choice perspective, it makes much more sense to judge a prospective partner’s moral character based on how she behaves towards in-network individuals than out-of-network ones.

Helping behaviors occurring within these interpersonal relationships are not part of the evidence that the MTT considers when it asks whether most people are compassionate, however. By relying upon studies of helping in zero-acquaintance scenarios, the MTT ends up evaluating compassion in contexts in which the relevant kinds of behavior are not naturally displayed, and compassionate motivation is most likely to be fragile. It is no wonder, then, that enhancers and inhibitors end up having such a strong affect upon the motivation to help, since it is understandably weak in these situations. In these contexts of evaluation, manifestations of compassion are indeed rare. But I argue that in more quotidian contexts – those that are most consistent with our ordinary moral character judgments – this would not the case.

Here, my argument yields a concrete prediction: we should expect to find much higher frequencies of compassion-relevant behaviors in contexts of existing social relationships than we find in contexts of zero-acquaintance. This prediction gains empirical support from evidence that we are more likely to empathically respond to the pain of in-network individuals than out-of-network individuals (Meyer et al., 2013); insofar as empathy is related to a range of prosocial behaviors (Batson et al., 1988), this suggests we are more likely to help people we know than strangers. To test this prediction, however, will require measures of compassionate behavior in everyday life, rather than in laboratory contexts (e.g. Bollich et al., 2016; Conner, Tennen, Fleeson, & Barrett, 2009). An additional prediction of the current view is that we should observe positive correlations between such behaviors and third-party judgments of compassion.

One objection to the current argument might be that there are good normative reasons for evaluating compassion in contexts of zero-acquaintance, and for holding people to a higher moral standard. For example, a Kantian might argue that these actions have greater moral worth and are more plausibly motivated by duty. Alternatively, it might be argued on consequentialist grounds that the identity of the target is morally irrelevant, and that all that matters is that a compassionate person reliably responds to the suffering of others. An Aristotelian might likewise attempt to argue from the unity-of-the-virtues thesis that a truly compassionate person must also be impartial or just. According to any of these moral frameworks, a person who does not meet the minimal thresholds for compassion contexts of zero acquaintance cannot be said to possess the virtue of compassion. If commonsense moral character judgments employ a weaker standard that focuses on behavior in the context of existing interpersonal relationships, then that just means that ordinary folk possess a normatively inadequate conception of compassion. Call this the *Strong Normative Standards Objection.*

The Strong Normative Standards Objection might be plausible own terms, but it cannot save the Character Error Theory as articulated by the MTT. This is because the two views offer very different kinds of error theories about ordinary moral character judgment. The CET does not hold that most people have the *wrong conception* of compassion; indeed, the MTT derives its own minimal thresholds for compassion from commonsense platitudes. Rather, the CET holds that, due to their lack of Mixed Trait concepts and the fundamental attribution error, ordinary folk systematically *misapply* their (otherwise correct) commonsense compassion concepts. People are not confused about what it takes to be compassionate, on this view. They just draw faulty inferences and form false beliefs about who is compassionate.

The Strong Normative Standards Objection, in contrast, holds that commonsense standards for compassion are themselves normatively inadequate. Ordinary folk are simply wrong about what it means to be compassionate. Their error lies in the content of their compassion concept, rather than the way in which they apply that concept to particular cases. This thesis is actually incompatible with the MTT, which appeals to commonsense intuitions to derive its minimal thresholds for compassion. According the Strong Normative Standards Objection, these commonsense ideas about compassion are, *ex hypothesi*, fundamentally confused.

More importantly for our purposes, the Strong Normative Standards Objection would actually be consistent with the central claim of this paper, which is that people regularly succeed in applying their ordinary moral character concepts in a consistent and reliable manner to track real psychological properties of individuals. Now, it might well be that this ability falls short of a true understanding of compassion as a virtue, which might also requires the appreciation of some unspecified philosophical considerations. This might be a legitimate avenue for critiquing ordinary moral character judgment (and I am certainly not arguing that ordinary conceptions of moral character are normatively correct), but it is a very different critique than the one advanced by the CET.[[8]](#footnote-8)

Like the MTT itself, the argument in this section has focused on the case of compassion. Does a similar argument hold in other domains where the MTT claims that virtues are rare? While a thorough survey is beyond the scope of this essay, one other domain of moral behavior does bear mentioning here: lying. Most of the empirical findings about lying discussed in Miller (2013) involve diary studies, in which participants self-reported on the lies that they told over the course of their day-to-day lives (e.g. DePaulo et al., 1998). While it has the downside of relying upon self-reports, this method avoids the most problematic feature of the helping studies: the social interactions in question spanned the participant’s entire social network, and did not just involve strangers. Overall, these studies tend to find that lying is relatively common, but that most of these lies are rated as less serious (DePaulo, 2004). Crucially, the vast majority of these less serious lies are told to casual acquaintances, rather than close friends or relatives. Moreover, lies told to close friends and family members tended to be viewed as much more serious, and were experienced as more distressing (DePaulo et al., 1998). This data suggests the tacit standards of honesty that people hold themselves to are indeed modulated by social scope: people think it is more dishonest to lie to more proximal members of one’s social network, and they adjust their honesty-relevant behavior accordingly.

**5. Conclusion**

Recent evidence from moral psychology suggests that moral character judgments are a pervasive feature of our everyday social cognition. The MTT raises a serious challenge to the epistemic reliability of these judgments. In response to this challenge, I have made the case that 1) we have good reason to believe that ordinary moral character judgments reliably track real psychological properties of individuals, and 2) that the evidence upon which the MTT is based is not representative of normal moral behavior. Thus, the challenge from the MTT is diffused.

Though situationism has not been the primary focus of this paper, it is worth noting that proponents of that view should also predict disagreement in our moral character judgments, given their views about the inconsistency of moral behavior. Situationism also draws support from experiments conducted in contexts of low or zero acquaintance. Thus, many of the arguments in this paper undermine the epistemic implications of situationism as well as those of the MTT. In general, the consistency with which ordinary folk agree in their moral character judgments will be problematic for any view that posits a mismatch between our moral trait concepts and the moral traits that people actually possess.

This discussion has also shed some light on the nature of everyday moral character judgment. When we judge someone we know to be virtuous or vicious, the standard of comparison we employ is typically tied to the set of social contexts that matter most to us in everyday life. By such standards, a person can count as compassionate or honest as long as they are compassionate or honest within their immediate social network. With respect to these standards, our moral character judgments appear to be quite reliable, and are even predictive of important life outcomes. However, it is likely that our judgments are less reliable when evaluating how far a person’s compassion extends beyond these social contexts. Perhaps some of the compassionate people we know might continue to reliably manifest their compassion even in contexts of zero-acquaintance; for other people, manifestations of compassion might be more fragile, and prone to the effects of enhancers and inhibitors like guilt, shame, and mood, as the MTT has convincingly argued. This only reveals the limits of ordinary moral character judgment, however; in the everyday contexts that matter most to us, our moral character judgments can be trusted to tell us what we need to know.[[9]](#footnote-9)

**References**

Albright, L., Dong, Q., Fang, X., Malloy, T. E., Kenny, D. A., Winquist, L., & Yu, D. (1997). Cross-Cultural Consensus in Personality Judgments. *Journal of Personality and Social Psychology*, *72*(3), 558–569.

Alfano, M. (n.d.). A plague on both your houses: Virtue theory after situationism and repligate. *Teoria*.

Alfano, M. (2011). Explaining Away Intuitions About Traits: Why Virtue Ethics Seems Plausible (Even if it Isn’t). *Review of Philosophy and Psychology*, *2*(1), 121–136.

Alfano, M. (2013). *Character as Moral Fiction*. Cambridge, UK: Cambridge University Press.

Allik, J., Realo, A., Mõttus, R., Borkenau, P., Kuppens, P., & Hřebíčková, M. (2010). How People See Others Is Different From How People See Themselves: A Replicable Pattern Across Cultures. *Journal of Personality and Social Psychology*, *99*(5), 870–882.

Anscombe, G. E. M. (1958). Modern moral philosophy. *Philosophy*, *33*(124), 1–19.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*(2), 150–166.

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A Short Measure of the Major Dimensions of Personality. *Journal of Personality Assessment*, *91*(4), 340–345.

Ashton, M. C., & Lee, K. (2010). Trait and source factors in HEXACO-PI-R self- and observer reports. *European Journal of Personality*, *24*(3), 278–289.

Ashton, M. C., Lee, K., de Vries, R. E., Perugini, M., Gnisci, A., & Sergi, I. (2006). The HEXACO model of personality structure and indigenous lexical personality dimensions in Italian, Dutch, and English. *Journal of Research in Personality*, *40*(6), 851–875.

Ashton, M. C., Lee, K., & Goldberg, L. R. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology*, *87*(5), 707–721.

Batson, C. D., Batson, J. G., Griffitt, C. A., Barrientos, S., Brandt, J. R., Sprengelmeyer, P., & Bayly, M. J. (1989). Negative-state relief and the empathy—altruism hypothesis. *Journal of Personality and Social Psychology*, *56*(6), 922.

Batson, C. D., Dyck, J., Brandt, J., Batson, J., Powell, A., & McMaster, M. (1988). Five studies testing two new egoistic alternatives to the empathy-altruism hypothesis. *Journal of Personality and Social Psychology1*, *55*(1), 52–77.

Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(1), 59–78.

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, *92*(1), 119–135.

Bollich, K. L., Doris, J. M., Vazire, S., Raison, C. L., Jackson, J. J., & Mehl, M. R. (2016). Eavesdropping on character: Assessing everyday moral behaviors. *Journal of Research in Personality*, *61*, 15–21.

Cann, A., & Blackwelder, J. G. (1984). Compliance and mood: A field investigation of the impact of embarrassment. *The Journal of Psychology*, *117*(2), 221–226.

Church, A. T. (2016). Personality traits across cultures. *Current Opinion in Psychology*, *8*, 22–30.

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2013). Agreement and similarity in self-other perceptions of moral character. *Journal of Research in Personality*, *47*(6), 816–830.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: a new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, *100*(5), 947–66.

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Connelly, B. S., & Ones, D. S. (2010). An Other Perspective on Personality: Meta-Analytic Integration of Observers’ Accuracy and Predictive Validity. *Psychological Bulletin*, *136*(6), 1092–1122.

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience Sampling Methods: A Modern Idiographic Approach to Personality Research. *Social and Personality Psychology Compass*, *3*(3), 292–313.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory Manual.* Odessa: Osychological Assessment Resources.

DePaulo, B. M. (2004). The many faces of lies. In A. G. Miller (Ed.), *The social psychology of good and evil* (pp. 303–326). New York: Guilford Press.

DePaulo, B. M., Kashy, D. A., Dahne, H., Epstein, J., Hairfield, J., Kirkendol, S., … Smith, C. (1998). Everyday lies in close and casual relationships, *74*(1), 63–79.

Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. Cambridge, UK: Cambridge University Press.

Dunbar, R. I. M. (2008). Cognitive Constraints on the Structure and Dynamics of Social Networks. *Group Dynamics*, *12*(1), 7–16.

Dunbar, R. I. M. (2014). The Social Brain: Psychological Underpinnings and Implications for the Structure of Organizations. *Current Directions in Psychological Science*, *23*(2), 109–114.

Dunlop, P. D., Morrison, D. L., Koenig, J., & Silcox, B. (2012). Comparing the Eysenck and HEXACO Models of Personality in the Prediction of Adult Delinquency. *European Journal of Personality*, *26*(3), 194–202.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, *37*(7), 1097.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2002). A Model of (Often Mixed Stereotype Content: Competence and Warmth Respectively Follow From Perceived Status and Competition. *Journal of Personality and Social Psychologyersonality and Social Psychology*, *82*(6), 878–902.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027.

Fleeson, W., Furr, R. M., Jayawickreme, E., Meindl, P., & Helzer, E. G. (2014). Character: The Prospects for a Personality-Based Perspective on Morality. *Social and Personality Psychology Compass*, *8*, 178–191.

Freeman, D., Aquino, K., & McFerran, B. (2009). Overcoming beneficiary race as an impediment to charitable donations: Social dominance orientation, the experience of moral elevation, and donation behavior. *Personality and Social Psychology Bulletin*, *35*(1), 72–84.

Funder, D. C. (1995). On the Accuracy of Personality Judgement: A Realistic Approach. *Psychological Review*, *102*(4), 652–670.

Furr, R. M. (2010). The Double-Entry Intraclass Correlation as an Index of Profile Similarity: Meaning, Limitations, and Alternatives. *Journal of Personality Assessment*, *92*(1), 1–15.

Gilbert, D. T., Malone, P. S., Aronson, J., Giesler, B., Higgins, T., Ross, L., … Trope, Y. (1995). The Correspondence Bias. *Psychological Bulletin*, *117*(1), 21–38.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–168.

Harman, G. (1999). Moral Philosophy Meets Social Psychology : Virtue Ethics and the Fundamental Attribution Error. *Proceedings of the Aristotelian Society, New Series*, *99*, 315–331.

Helgeson, C. (2013). The Confirmational Significance of Agreeing Measurements. *Philosophy of Science*, *80*(5), 721–732.

Helzer, E. G., Furr, R. M., Hawkins, A., Barranti, M., Blackie, L. E. R., & Fleeson, W. (2014). Agreement on the Perception of Moral Character. *Personality and Social Psychology Bulletin*, *40*(12), 1698–1710.

Hilbig, B. E., & Zettler, I. (2009). Pillars of cooperation: Honesty-Humility, social value orientations, and economic behavior. *Journal of Research in Personality*, *43*(3), 516–519.

Jayawickreme, E., Meindl, P., Helzer, E. G., Furr, R. M., & Fleeson, W. (2014). Virtuous states and virtuous traits: How the empirical evidence regarding the existence of broad traits saves virtue ethics from the situationist critique. *Theory and Research in Education*, *12*(3), 283–308.

Kamtekar, R. (2004). Situationism and Virtue Ethics on the Content of Our Character. *Ethics*, *114*(3), 458–491.

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, *98*(2), 155–163.

Landy, J. F., & Uhlmann, E. L. (2018). Morality is personal. In K. Gray & J. Graham (Eds.), *Atlas of Moral Psychology* (pp. 121–132). New York, NY: Guilford Publications.

Latané, B., & Rodin, J. (1969). A lady in distress: Inhibiting effects of friends and strangers on bystander intervention. *Journal of Experimental Social Psychology*, *5*(2), 189–202.

Lee, K., & Ashton, M. C. (2008). The HEXACO personality factors in the indigenous personality lexicons of English and 11 other languages. *Journal of Personality*, *76*(5), 1001–1054.

Lee, K., & Ashton, M. C. (2017). Acquaintanceship and self/observer agreement in personality judgment. *Journal of Research in Personality*, *70*, 1–5.

Lee, K., Ashton, M. C., & De Vries, R. E. (2005, April). Predicting workplace delinquency and integrity with the HEXACO and Five-factor models of personality structure. *Human Performance*. Lawrence Erlbaum Associates, Inc.

Lee, K., Ashton, M. C., Pozzebon, J. A., Visser, B. A., Bourdage, J. S., & Ogunfowora, B. (2009). Similarity and Assumed Similarity in Personality Reports of Well-Acquainted Persons. *Journal of Personality and Social Psychology*, *96*(2), 460–472.

Luan, Z., Poorthuis, A. M. G., Hutteman, R., Denissen, J. J. A., Asendorpf, J. B., & van Aken, M. A. G. (2018). Unique predictive power of other-rated personality: An 18-year longitudinal study. *Journal of Personality*.

Martin, J. W., & Cushman, F. (2015). To punish or to leave: Distinct cognitive processes underlie partner control and partner choice behaviors. *PLoS ONE*, *10*(4), 9–14.

Meyer, M. L., Masten, C. L., Ma, Y., Wang, C., Shi, Z., Eisenberger, N. I., & Han, S. (2013). Empathy for the social suffering of friends and strangers recruits distinct patterns of brain activation. *Social Cognitive and Affective Neuroscience*, *8*(4), 446–454.

Miller, C. B. (2013). *Moral character: An empirical theory*. Oxford: Oxford University Press.

Miller, C. B. (2014). *Character and moral psychology*. Oxford: Oxford University Press.

Miller, C. B. (2017). Character and Situationism: New Directions. *Ethical Theory and Moral Practice*, *20*(3), 459–471.

Mischel, W. (2004). Toward an Integrative Science of the Person. *Annual Review of Psychology*, *55*(1), 1–22.

Mischel, W., & Shoda, Y. (1995). A cognitive-active system theory of personality: Reconceptualising situations, dispositions, dynamics and invariance in personality structure. *Psychological Review*, *12*(1), 246–268.

Ott-Holland, C. J., Huang, J. L., Ryan, A. M., Elizondo, F., & Wadlington, P. L. (2014). The effects of culture and gender on perceived self-other similarity in personality. *Journal of Research in Personality*, *53*, 13–21.

Paris, P. (2016). Scepticism about Virtue and the Five-Factor Model of Personality. *Utilitas*, *29*(4), 1–30.

Powell, J., Lewis, P. A., Roberts, N., García-Fiñana, M., & Dunbar, R. I. M. (2012). Orbital prefrontal cortex volume predicts social network size: An imaging study of individual differences in humans. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1736), 2157–2162.

Regan, D. T., Williams, M., & Sparling, S. (1972). Voluntary expiation of guilt: A field experiment. *Journal of Personality and Social Psychology*, *24*(1), 42.

Ross, L. (1977). The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process. *Advances in Experimental Social Psychology*.

Roth, M., & Altmann, T. (2019). A multi-informant study of the influence of targets’ and perceivers’ social desirability on self-other agreement in ratings of the HEXACO personality dimensions. *Journal of Research in Personality*, *78*, 138–147.

Russell, D. C. (2009). *Practical intelligence and the virtues*. Oxford University Press.

Sabini, J., & Silver, M. (2005). Lack of Character? Situationism Critiqued. *Ethics*, *115*(3), 535–562.

Snow, N. E. (2010). *Virtue as social intelligence: An empirically grounded theory*. New York, NY: Routledge.

Sripada, C. S., & Konrath, S. (2011). Telling more than we can know about intentional action. *Mind & Language*, *26*(3), 353–380.

Stiller, J., & Dunbar, R. I. M. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, *29*(1), 93–104.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159–171.

Sutcliffe, A. G., Dunbar, R. I. M., Binder, J. F., & Arrow, H. (2012). Relationships and the social brain: Integrating psychological and evolutionary perspectives. *British Journal of Psychology*, *103*, 149–168.

Thielmann, I., Zimmermann, J., Leising, D., & Hilbig, B. E. (2017). Seeing is Knowing: On the Predictive Accuracy of Self- and Informant Reports for Prosocial and Moral Behaviours. *European Journal of Personality*, *31*(4), 404–418.

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A Person-Centered Approach to Moral Judgment. *Perspectives on Psychological Science*, *10*(1), 72–81.

Uhlmann, E. L., Zhu, L. L., & Diermeier, D. (2014). When actions speak volumes: The role of inferences about moral character in outrage over racial bigotry. *European Journal of Social Psychology*, *44*(1), 23–29.

Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science*, *20*(2), 104–108.

Vazire, S., Chung, C., Freeman, H., Mehta, P., Baquero, C., Harrison, H., … Beard, S. (2010). Who Knows What About a Person ? The Self – Other Knowledge Asymmetry ( SOKA ) Model. *Journal of Personality and Social Psychology*, *98*(2), 281–300.

Westra, E. (2019). Getting to know you: accuracy and error in judgments of character. *Mind & Language*.

Whewell, W. (1989). Novum organon renovatum. In R. E. Butts (Ed.), *In William Whewell: Theory of Scientific Method,*. Indianapolis: Hackett.

1. Of course, all things are not always equal, and so individual behavior is always quite variable from one situation to the next. Even highly extraverted people will, over the course of a week, display some behaviors consistent with extreme introversion. But this is consistent with displaying high average levels of extraversion, which remain stable over time (Epstein, 1979; Fleeson, 2001). [↑](#footnote-ref-1)
2. Note that this is merely one of many minimal thresholds for compassion that Miller articulates. [↑](#footnote-ref-2)
3. One general worry here is that many of the studies Miller draws upon would be considered statistically underpowered and vulnerable to non-replication by contemporary standards, which have changed rapidly since the advent of the replication crisis in psychology (Alfano, n.d.; Collaboration, 2015; C. B. Miller, 2017). In my view, these considerations should weaken our confidence in some of the results Miller relies on, but not erase it entirely. His survey of the literature is broad enough that his overall argument could still hold even if it were found that a significant proportion of these studies do not replicate. [↑](#footnote-ref-3)
4. The “fundamental attribution error” is our tendency to explain behaviors in terms of dispositional causes even when explanations involving situational causes are more plausible (Gilbert et al., 1995; Ross, 1977). Notably, Miller expresses significant unease with the way these findings are interpreted (C. B. Miller, 2014, pp. 158–170). Elsewhere, I have argued that the fundamental attribution error and other biases like it are less detrimental to the reliability of our character judgments than they seem, because they do not reflect the way that our character judgments update in response to new information (Westra, 2019). In what follows, I focus mainly on Miller’s claims about our trait concepts and set the issue of the fundamental attribution error to one side. [↑](#footnote-ref-4)
5. The cross-cultural generalizability of this research is a complex matter. Some patterns of agreement in personality judgment, such as differences in how one rates the self and how one rates others, have been found to replicate across numerous cultural samples (Albright et al., 1997; Allik et al., 2010). However, other aspects of personality judgment, such as the degree to which people tend to view themselves as similar to others, vary based on several cultural factors, such as institutional collectivism (Ott-Holland, Huang, Ryan, Elizondo, & Wadlington, 2014). An added source of complexity in these comparisons is that it is not clear whether the same basic personality dimensions emerge in all cultures (Church, 2016); if personality traits themselves are different in some cultures, this limits our ability to assess agreement using the same standardized measures of personality. [↑](#footnote-ref-5)
6. Miller (2014, p. 144) acknowledges that the content of the Honesty-Humility factor is distinctively moral, but denies that scoring highly on this factor would be enough to count as virtuous according to his various minimal thresholds. This dismissal is too hasty, however. First, even if we accepted Miller’s claim that high scores on Honesty-Humility would not suffice for genuine virtue, the corollary claim that low scores on this dimension would not suffice for *vice* is quite implausible. Second, answers to the Honesty-Humility questionnaire items correlate strongly with how people rate themselves using adjectives that express both virtue and vice concepts, suggesting that these questions do reflect commonsense standards for how these terms are used (Ashton et al., 2006). [↑](#footnote-ref-6)
7. Notably, Helzer et al. (2014) was conducted on both a community sample and a student sample. Agreement about moral character traits was significantly higher in the community sample than the student sample, which accounts for some of the weakness in the trait-level agreement. The authors suggest that this may have been due to the fact that informants in this group tended to use a more restricted range of the trait scales than the community sample, indicating that they refrained from making more negative trait attributions about their peers, or were otherwise unable to identify less moral targets. Alternatively, they suggest that character traits may simply be less consistent in younger populations when it is still developing, and that moral character does not stabilize until later in adulthood (Helzer et al., 2014, pp. 1707–1708). [↑](#footnote-ref-7)
8. Suppose, by analogy, that deontology was the correct moral theory, but that ordinary folk intuitions about morality were purely consequentialist. The deontologist might readily concede that ordinary folk intuitions actually succeed in reliably tracking consequences, while still denying that ordinary folk have the correct conception of morality. Their quarrel with the folk would not about their capacity for reliable consequentialist judgment, but rather a normative disagreement. This case is like the Strong Normative Standards Objection. In contrast, one might accept that the folk are right and that consequentialism is correct, while also holding that ordinary intuitions about consequences are highly unreliable in practice. This case is more like the CET. [↑](#footnote-ref-8)
9. I am grateful to Peter Carruthers, John Michael, and Jennifer Nagel for comments on drafts of this paper. [↑](#footnote-ref-9)