

The Prospects and Challenges of Measuring a Person's Overall Moral Goodness

Jessie Sun¹ and Eric Schwitzgebel²

¹Washington University in St. Louis

²University of California, Riverside

Author Note

Correspondence concerning this article should be addressed to Jessie Sun, Department of Psychological and Brain Sciences, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130. Email: jessie.sun@wustl.edu.

Version: Revised manuscript submitted for publication (May 29, 2025)

Academic Abstract

This paper integrates psychological and philosophical theory and research to explain what it would take to measure a person's general morality (i.e., their overall moral goodness). We outline the conceptual requirements for measuring general morality and argue that it would be difficult to operationalize morality in a way that satisfies these requirements. Self-report, informant report, behavioral, and biological measurement approaches also have substantial methodological limitations. These conceptual and methodological challenges limit the validity of measures of general morality more than they do for most other psychological traits. This exercise yields general lessons about the challenges that also confront the design of narrower, less ambitious morality measures (e.g., of specific moral virtues). Some—but not all—of these challenges can be mitigated when the measurement aims are more modest. It is therefore important to be transparent and intellectually humble about what we can and cannot conclude based on various moral assessments.

Keywords: moral measurement, moral assessment, general morality, overall morality, moral character, moral psychology

Public Abstract

It is possible to measure how morally good or bad a person is, overall? We argue that general morality (i.e., a person's overall moral goodness) is much more difficult to measure than many other psychological traits. Conceptually, it is not entirely clear that there is such a trait as overall moral goodness that applies consistently across people, cultures, and contexts, much less that either researchers or the people who are being measured could agree on its elements. Practically, any measure would need to rely on self-reports, reports by others, direct behavioral observations, or biological measures. There are serious problems (e.g., bias, incompleteness, infeasibility) with each of these strategies. Similar difficulties arise, though less severely, in measuring more specific moral tendencies (e.g., honesty, kindness, and moral values). It is therefore important to be humble about what we can and cannot conclude when attempting to measure a person's overall moral goodness.

The Prospects and Challenges of Measuring a Person's Overall Moral Goodness

Is it possible to measure a person's *general morality*; that is, how morally good (vs. bad) they are, overall, compared to other people? Imagine what power we'd have if a general "moralometer" existed. It could help inform decisions about who we should befriend, marry, hire, lend large sums of money to, trust with our darkest secrets, vote into positions of power, and assiduously avoid. An accurate morality measure would also help us resolve philosophical debates about whether moral people tend to be happier (Sun et al., 2025), understand whether society is making moral progress (Mastroianni & Gilbert, 2023; Pinker, 2018), and assess whether contemplative practices, humanities education, exposure to literary fiction, or a career studying ethics tends to improve people's morality (Berryman et al., 2023; Christiansen, 2023; Nussbaum, 1997; Schwitzgebel & Rust, 2014).

We start from the assumption that people differ meaningfully from one another in how they perceive, think about, and respond to moral issues (Sun & Smillie, 2024), and in the positive and negative consequences they have on the world. Historically, situationists were skeptical of the existence of stable individual differences—moral or otherwise (Doris, 2002; Mischel, 1968)—and personality psychologists also shied away from the explicit study of morality for being "nonscientific" (for reviews, see McAdams & Mayukha, 2024; Sun & Smillie, 2024). Today, a growing body of empirical evidence demonstrates that moral personality is a viable topic of scientific inquiry (Fleeson et al., 2014). Individual differences in various aspects of moral functioning (e.g., values, behaviors, virtues, beliefs, emotions, reasoning) exist and predict meaningful social, political, workplace, and well-being outcomes (e.g., Atari et al., 2023; Cohen et al., 2012; Collier-Spruel et al., 2019; Sun et al., 2025). However, here, we argue that it would

be uniquely difficult to develop a general moralometer that could accurately quantify a person's overall moral goodness.

From a character-focused perspective, a person's overall moral goodness reflects the overall balance of various virtues or vices (e.g., honesty, compassion, fairness, loyalty, purity)—in other words, their “global moral character” (Furr et al., 2022). However, our conclusions do not assume a virtue ethics or character psychology framework. For example, a consequentialist might conceptualize a person's moral goodness in terms of their overall contribution to the sum total of good minus bad in the world, and a deontologist might focus on a person's overall tendency to abide by moral rules and fulfill their obligations. For this reason, in this article, we use the terms “general morality” or “overall moral goodness” (which we use interchangeably), rather than “moral character” (which has been the focus of previous reviews on the topic of individual differences in moral goodness; see below).

Pioneering reviews on moral character have alluded to the idea that it is challenging to reach consensus on what “morality” is (Cohen & Morse, 2014; Fleeson et al., 2014; Fleeson et al., 2015)—but have not explained why this is, nor thoroughly evaluated the merits of different methodological approaches for measuring moral character (for a brief review, see Wright et al., 2020). Instead, existing reviews have broadly defined moral character as “an individual's disposition to think, feel, and behave in an ethical versus unethical manner, or as the subset of individual differences relevant to morality” (Cohen & Morse, 2014, p. 45), or “characteristics that are descriptive of actions, cognitions, emotions, and motivations that are considered to be relevant to right and wrong according to a relevant moral standard” (Fleeson et al., 2014, p. 181). Such definitions have left the question of what constitutes moral goodness open to conceptual debate and individual researchers' predilections. This ecumenical approach has allowed research

on moral personality to proceed in the face of eclectic definitions of morality (Fleeson et al., 2015) and a corresponding variety of morality measures.

By contrast, in this paper, we grapple with—rather than set aside—the conceptual and methodological challenges that would limit the validity of a general moralometer; that is, the degree to which scores from a moralometer can be interpreted as reflecting a person’s overall moral goodness. Specifically, we ask: What would it take to measure a person’s actual level of overall moral goodness (compared to other people) in a way that is reasonably consistent across sociocultural contexts, and can be feasibly implemented using existing methods of psychological assessment? Combining our interdisciplinary perspectives from philosophy and psychology (as recommended by Fleeson et al., 2014), we will detail the conceptual difficulties involved quantifying overall moral goodness (see Table 1) and the practical challenges in capturing the relevant thoughts, feelings, motivations, and behaviors using available methods for psychological assessment (see Table 2). These challenges would substantially limit the validity of a general moralometer. We also contend that these conceptual and methodological challenges arise more severely for most moral traits than for most non-moral traits (see Discussion for a comparison with extraversion and well-being).

To some readers, it might already seem obvious that an accurate, general purpose “moralometer” could never be built. Our project is to think systematically about the conceptual and methodological challenges that would limit the validity of a general moralometer. Furthermore, this exercise yields general lessons about the challenges that also confront the design of narrower, less ambitious morality measures.

Despite these various challenges, we are not complete pessimists about the prospects of measuring any aspect of a person’s moral goodness, especially for limited scientific purposes.

Borsboom and colleagues (2004) consider a test to be valid for measuring an attribute if “(a) the attribute exists and (b) variations in the attribute causally produce variation in the measurement outcomes” (p. 1061). However, validity is often conceptualized as a matter of degree (AERA et al., 2014; Cronbach & Meehl, 1955; Messick, 1998). *The Standards for Educational and Psychological Testing* further emphasizes interpretations and intended uses, defining validity as “the degree to which evidence and theory support the interpretations of test scores for proposed uses” of a test (AERA et al., 2014, p. 11). If validity is understood to be a matter of degree and dependent on what we intend to do with the results, the relevant question will always be whether any purported measure of general morality is valid *enough* for the purposes at hand and whether its shortcomings are manageable and well-understood (see our Recommendations section in the Discussion).

We will also discuss how some of the conceptual and methodological challenges of measuring overall morality can be mitigated when the measurement aims are more modest (e.g., specific operationalizations, behaviors, or contexts), when research aims and conclusions are appropriately calibrated to the strengths of the measure (e.g., if self-report measures are interpreted as capturing moral *self-perceptions*; see Table 3), and when targeting narrower moral traits. To illustrate this in more detail, in a parallel discussion in the Supplemental Material (Section 3), we evaluate the extent to which our conceptual and methodological critiques apply to two examples of more specific moral traits—compassion and honesty. This comparison demonstrates that although serious conceptual and methodological difficulties still plague the measurement of even specific moral traits, they are not as severe as for overall morality. Finally, we conclude with recommendations (see Table 4) for increasing transparency, intellectual humility, and ethical caution in moral measurement.

Conceptual Requirements

To measure overall moral goodness, we must have a substantive theory of what it is (Briggs et al., forthcoming). In the broadest sense, actions—and maybe also thoughts, feelings, and motivations—are morally good to the extent they are ethically virtuous, admirable, praiseworthy, or right. Paradigmatic examples include fulfilling one’s promises and obligations (absent compelling reason not to); aiding others (if doing so won’t cause more harm than it prevents); avoiding needless cruelty; and carefully, informedly, and respectfully trying to bring about more good than bad in the world. A person’s overall moral goodness would then involve some summation or balancing of all their moral actions, traits, tendencies, and other morally relevant features, good and bad. Our characterization is intentionally vague, given the extensive debate over what exactly is ethically virtuous, admirable, praiseworthy, and right (as we discuss further below).

A valid general moralometer would permit comparisons by which pairs of individuals (across some broad range of people) can be accurately characterized as more or less moral than each other. Following Borsboom and colleagues’ (2004) definition of test validity, a moralometer is valid for measuring general morality if (a) general morality exists and if (b) variations in general morality causally produce variations in scores on the moralometer. Here, we propose four conceptual requirements for constructing a sufficiently valid general moralometer: (1) There must be general facts about a person’s overall moral goodness; (2) The measure must correctly identify which characteristics are morally good or bad; (3) The measure must correctly weigh different components of morality against each other; and (4) The measure must apply clearly and consistently across people and time. Conceptual Requirement 1 pertains to the existence of general morality, whereas Conceptual Requirements 2–4 and the Methodological

Requirements (outlined in the next section) pertain to factors that affect the degree to which variations in general morality produce variations in scores on the moralometer. Together, these proposed requirements constitute serious obstacles for any measure that claims to accurately capture a person's overall moral goodness.

We begin by introducing two broad approaches to constructing a moralometer on the basis of *flexible* or *fixed* criteria. We then consider the extent to which flexible and fixed measures plausibly satisfy each of the four conceptual requirements for general morality (see Table 1).

Flexible or Fixed Measures

To construct a general measure of a person's overall moral goodness, we can use either (a) *flexible* criteria based on judges' understandings of how to evaluate and weight the various facets of morality into a general score, or (b) *fixed* criteria that deliver a general score based on criteria selected by and weighted by the researchers. *Mixed* criteria might have both fixed and flexible aspects—for example, creating a weighted average of one flexible participant-determined assessment and several fixed, researcher selected measures, or weighting four different facets of morality equally but permitting judges to reach flexible judgments about each facet individually.

The simplest flexible measure might be a general single-item question about a person's overall morality: "How moral are you?" (self-report) or "How moral is [target's name]?" (informant report), and the response scale might range from 0 (*completely immoral*) to 100 (*morally perfect*). The key feature of a flexible measure is that it does not specify which traits or behaviors are "good" or "virtuous." Such a measure defers all questions about the substance of morality to judges' personal understandings—even if the judge's understandings differ from the

target's. (Throughout the article, we use the term "target" to refer to the person whose morality is being measured, and the term "judge" to refer to any person who is providing an evaluation of a target's morality. In the case of self-reports, the target and the judge are the same person.)

Table 1*Conceptual Requirements for Constructing Flexible vs. Fixed Measures of General Morality*

Conceptual requirement	Flexible	Fixed
1. There must be general moral facts about a person's overall morality.		
Realism: There must be facts about what is (im)moral; <i>and</i>	✓	✓
Universalism: The same general things must be (im)moral for different people or in different groups; <i>and</i>	!	!!
Generalism: What is (im)moral must not depend on highly particular features of specific situations.	!	!!
2. The measure must correctly identify which characteristics are morally good or bad.		
Judges' idiosyncratic moral judgments are correct; <i>or</i>	!!	-
Commonsense ethics, as operationalized by the researchers, is correct; <i>or</i>	-	!!
The favored expert ethical framework is correct.	-	!!
3. The measure must correctly weigh different components of morality against each other.		
Unity: General morality must be a coherent construct; <i>and</i>	!!	!!
Commensurability: There must be a common "currency" in which components of (im)morality can be appropriately compared.	!!	!!
4. The measure must apply clearly and consistently across people, groups, and time.		
Transparency: It must be clear what is being measured; <i>and</i>	!!	✓
Equivalence:		
A fixed measure must capture the same thoughts, feelings, and/or behaviors across people; <i>and</i>	-	✓
A fixed measure must have the same moral significance across people; <i>or</i>	-	!!
For flexible measures, judges must use the same criteria to judge a person's (im)morality; <i>and, depending on the aim</i>	!!	-
The measure must be psychologically relevant.	!	!!

Note. - = Not applicable; ✓ = Requirement is likely satisfied; ! = Significant difficulty; !! = Major difficulty. See the Supplemental Material for an extension of this framework to compassion and honesty (see Table S2) and extraversion and well-being (see Table S4).

An alternative to the flexible, subjective approach is to use a fixed set of criteria, which could be derived from philosophical theories, religious prescriptions, cultural values, or a consensus-based commonsense morality. For example, one might operationalize general morality as a composite of moral virtues (e.g., compassion, honesty, fairness, and loyalty; Furr et al., 2022; Helzer et al., 2014); as a composite of traits such as honesty-humility, guilt-proneness, and moral identity (Cohen et al., 2014; Helzer et al., 2024); as the extent to which a person embodies utilitarian values; or as a composite based on a person's meat-eating, environmental, and charitable donation behaviors. This approach is *fixed* in the sense that a person's moral temperature is assessed based on the extent to which they align with prespecified criteria—even if that person thinks that such tendencies are morally irrelevant or are even indicative of *immorality*.

Are There General Moral Facts?

The existence of individual differences in general morality depends on the existence of general moral facts. Moral relativists, particularists, and skeptics challenge the idea that such facts exist. *Moral relativists* hold that different things are (im)moral for different people or in different groups or cultures (Gowans, 2021; Wong, 2006). This isn't the obvious point that people sometimes differ substantially in what they *judge* to be (im)moral (discussed below) but the less obvious idea that what *actually is* (im)moral differs substantially between people or groups. *Moral particularists* believe that morality is not about applying consistent rules; rather, what's morally right or wrong frequently depends on particular features of specific situations that cannot be fully codified in advance (Dancy, 2017). *Moral skeptics* (Sinnott-Armstrong, 2019) reject the validity of moral claims altogether, viewpoint-relativized or not, holding that there are no moral facts.

If the most extreme forms of moral relativism, particularism, or skepticism are true, then no moralometer could possibly work because there won't be general truths about people's overall moral goodness, or the truths will be so situation-dependent as to defy any practical attempt at measurement. We assume the falsity of the most extreme forms of moral relativism, particularism, and skepticism according to which genocide, for example, doesn't warrant universal condemnation. However, moderate versions of moral relativism and particularism cannot be so easily discounted. Cultures might reasonably differ, for example, on the age of sexual consent (e.g., 16 in Cuba vs. 18 in Turkey), and social groups might reasonably differ in standards of generosity in sharing resources with neighbors and kin. There might be no culture-independent fact of the matter about which set of norms is more correct. And, in accord with moderate particularism, how morally bad it is to fail to help a stranger or how morally good it is to donate a \$100 windfall can vary substantially depending on specific details that are difficult to know from afar or codify in advance (e.g., maybe helping a male stranger is riskier for women, maybe the \$100 would help cover necessary dental work for one's children).

To the extent that moderate versions of moral relativism or particularism are correct, they introduce substantial challenges. Moderate moral relativism and particularism suggest that the farther the researcher or judge stands from the social group or situation of the target, the more likely they are to apply inappropriate standards. Because fixed measures are applied equally to all targets, such measures will be, at best, incomplete or noisy, failing to capture important group- or situation-specific aspects of morality. At worst, fixed measures risk inappropriately importing researchers' own group-specific standards. If so, flexible measures, employed by knowledgeable judges who are sensitive to locally appropriate standards, will have an advantage. However, even flexible measures will fail if they are employed by outgroup judges who employ

the wrong standards or are ignorant of relevant situational details. If moral skepticism is correct, there are no facts about whether someone “really” is or is not moral; but a measure that is good enough in other respects might still capture something general about a person’s character relative to an assumed moral viewpoint (an issue we return to in the Discussion).

One concern that we set aside here is whether general morality is a reflective latent variable (Edwards & Bagozzi, 2000); that is, an underlying moral disposition that causes various observed indicators such as trait levels of compassion, honesty, and fairness. An alternative to the reflective model is a formative model which reverses the causal direction between a variable and its indicators (Edwards & Bagozzi, 2000). For example, socioeconomic status (SES) is typically assessed as a composite of indicators such as income, education, occupational prestige, and wealth (Antonoplis, 2023). If SES is defined as the total amount of valued social and economic resources (Antonoplis, 2023), then a change in any of these indicators causes changes in SES (consistent with a formative model), whereas a change in overall SES need not change all of its indicators (contra the key assumption of a reflective model). General morality could likewise be a formative variable that simply summarizes its indicators without causing them. If so, in principle, purity, fairness, loyalty, compassion, and other traits or behaviors might belong to general morality without correlating at all with one another. A utilitarian consequentialist might, for example, be interested in measuring the sum total of the good and bad consequences of a person’s behavior without psychological assumptions about the intercorrelations among these indicators or the predictive validity of the composite—though for some caveats about the usefulness of this, see our discussion of “psychological moral relevance” below. Though we reject the strongest forms of psychological situationism—which deny the existence of robust and stable personality traits (Doris, 2002; Kenrick & Funder, 1988; Lucas & Donnellan, 2009;

Mischel et al., 1968; Ross & Nisbett, 1991)—even if situationism were true, general morality could still exist as a formative construct.

How Should We Determine What is Morally Good vs. Bad?

Assuming that some general facts exist about what is morally good or bad, we need a measure that accurately captures these facts. Previous reviews of personality-based approaches to morality have refrained from providing guidance on how to determine what is morally good vs. bad (Cohen & Morse, 2014; Fleeson et al., 2014; Fleeson et al., 2015). We address this gap by reviewing the strengths and limitations of three main options: Should we rely on the moral understanding of the individual, the consensus of many people, or a formal moral framework?

Idiosyncratic Morality. Flexible measures assume that the self or informants (“judges”) know the relevant moral facts. This assumption faces the problems of *moral disagreement* (people disagree about what is morally good vs. bad) and *moral ignorance* (even if people agree with each other, they might all be mistaken). Moral disagreement and ignorance pose substantial challenges for measurement equivalence and psychological relevance (issues we discuss below). Here, however, we focus on their implications for the accuracy of flexible moral judgements.

There is clearly some degree of convergence among people on what constitutes moral goodness, both within and between cultures. For example, within U.S. culture, when asked to rate the moral relevance of various personality traits, undergraduate students from two different universities came to almost identical conclusions ($r = .98$) about which traits were more (e.g., compassion, honesty, fairness) vs. less (e.g., anxiety, sociability, creativity) morally relevant (Sun & Goodwin, 2020; see also Fleeson et al., 2024). Between cultures, ancient Chinese philosophers, for example, celebrate familiar virtues and behaviors, such as “trustworthiness” (*xìn 信*), “loyalty” (*zhōng 忠*), and “benevolence” (*rén 仁*) (Van Norden & Ivanhoe, 2023).

Nevertheless, there is also considerable diversity in people's moral judgments (Awad et al., 2020; Graham et al., 2016; Meindl & Graham, 2014). For example, political liberals care relatively more about the moral foundations of harm and equality (compared to conservatives), whereas conservatives care relatively more about loyalty, purity, respect for authority, and proportionality (compared to liberals; Atari et al., 2023). Moral judgments of the same behavior can even go in the opposite direction: Pro-life and pro-choice activists may both judge themselves to be highly moral when enacting their moral values but would judge each other to be acting immorally. Even worse, people with noxious worldviews might ignorantly rate themselves and their counterparts very positively (e.g., notorious Nazi Adolf Eichmann appears to have thought highly of fellow Nazis' moral character; Arendt, 1963). Beyond broad differences in worldview, individual judges might have unusual, incompatible, or erroneous moral views. Such idiosyncrasies would introduce substantial noise and possibly also substantial systematic error if the goal is to measure who is in fact more vs. less morally good. (Note that this problem is mitigated if the ratings are interpreted not as revealing facts about a target's moral goodness but as revealing judges' subjective perceptions of a target's moral goodness; we expand on the value of more modest research goals such as this in the Discussion.)

Commonsense Morality. One alternative to relying on idiosyncratic judgments is to defer to what people typically agree upon as being moral or immoral (Fleeson et al., 2014). That is, it is possible to use a consensus-based approach to develop a fixed measure of a person's general morality. Consensus ratings can then be used to select traits for inclusion in a composite measure. For example, U.S.-based participants in Sun and colleagues (2025) generally rated traits such as compassion, honesty, fairness, and loyalty as being highly morally relevant. Hadza hunter-gatherers generally agree that effort and generosity are relevant to moral character (Smith

& Apicella, 2020). Delphi, an A.I. system that generates moral judgments about a variety of everyday situations, was trained on “commonsense norm bank” that comprised commonsense moral judgments of 1.7 million everyday situations (Jiang et al., 2022). Commonsense morality could also be inferred from people’s perceptions of and conformity to behavioral norms (e.g., concerning littering, hygiene, bullying, or sexuality; Bicchieri, 2017; Cialdini, 2007; Lindström et al., 2018; Paluck & Shepherd, 2012).

Such an approach treats commonsense morality as approximately correct. However, institutionalized slavery and the disenfranchisement of women were once widely considered morally acceptable, and some would argue that commonsense morality similarly fails to recognize currently ongoing moral catastrophes (Enriquez, 2021; Williams, 2015). Conversely, some practices that were once widely condemned (e.g., homosexuality, having children out of wedlock) are now generally considered to be morally acceptable (at least in Western, liberal, secular cultures). More generally, human cooperative tendencies evolved in small communities in which it was arguably adaptive to have a bias towards the near future while discounting outcomes for future generations (Law et al., 2023; Syropoulos et al., 2023), to restrict one’s altruism to a small circle of kin (Crimston et al., 2016), to be slow to sympathize with larger numbers of people (Jenni & Loewenstein, 1997), and to weigh harmful actions more than harmful omissions (Spranca et al., 1991). Some philosophers would argue that such moral dispositions aggravate the greatest moral challenges of today’s modern world (Caviola et al., 2021; Greene, 2014; Jaeger & van Vugt, 2022; Persson & Savulescu, 2012). If ordinary moral character judgments derive mostly from everyday social interactions (Westra, 2022) rather than the extent to which a person is doing good for the world more broadly, then on some philosophical conceptions of morality, ordinary judgments will omit a large part of morality.

Similarly, if ordinary moral judgments are substantially influenced by luck in outcome (e.g., regarding attempted murder as less bad than actual murder), or if they regard omissions as less bad than actions (e.g., passively letting someone be harmed as less bad than actively harming them), then on philosophical conceptions that treat luck or the action/omission difference as morally irrelevant, commonsense moral standards will be systematically mistaken (Kagan, 1989; Zimmerman, 2015). For these reasons, treating shared, non-expert intuitions as a reliable guide to moral truth is scientifically and philosophically problematic.

Ethical Frameworks. If researchers wish to circumvent the major issues associated with a reliance on participants' idiosyncratic moral judgments or commonsense morality, they must make—and be prepared to defend—explicit moral commitments as to what a more appropriate fixed standard is. This is a big ask, given that even philosophers and theologians who spend their lifetimes reflecting on such issues do not agree with one another on what the moral truths are (Shafer-Landau, 1994; Yaden & Anderson, 2021). If researchers have an incorrect moral theory, their moral measurements will likely also be misguided.

For example, is being a moral person about consistently making *consequentialist* decisions that maximize overall benefits and minimize overall harms (Sinnott-Armstrong, 2022), about following *deontological* rules or duties (e.g., do not lie; do not kill; Alexander & Moore, 2021), or about embodying particular *virtues* (e.g., honesty, compassion, temperance; Hursthouse & Pettigrove, 2023)? These ethical frameworks often produce different conclusions about moral goodness. The notorious “trolley problem” case of killing one innocent person to save five others—endorsed by (most) consequentialists but condemned by (most) deontologists—is only the beginning. Complicating the issue, advocates of each framework often disagree about mid-level principles or particular actions. For example, how much should we consider the

consequences for currently existing lives vs. future generations, and is it better to maximize total happiness or average happiness (Caviola et al., 2022; Parfit, 2011)? Which particular deontological rules or virtues are morally relevant? Decisions among general frameworks—and among their variants—could have huge ramifications for moral measurement.

How Should We Weight Different Aspects of Morality?

Goods are considered “incommensurable” if there is no fact of the matter about how they should be weighed relative to each other. \$20 bills and \$10 bills are commensurable: One of the former is worth exactly two of the latter. In contrast, it is not clear that there is a common “moral currency” upon which moral and immoral characteristics can be readily compared (Chang, 2015). For example, who is more moral: Tara (who is stingy but fair) vs. Nicholle (who is generous but plays favorites); Justin (who regularly helps others for selfish reasons) vs. Ryan (who has morally pure intentions but rarely follows through); or Anika (who kills a dog) vs. Nathan (who fails to prevent the death of two dogs)? And how should people’s conduct across the civic, professional, and personal spheres—in which people may have different relational obligations (Earp et al., 2021; Rai & Fiske, 2011)—be weighted? There might be no conceptually rigorous way to generate a single numerical score that combines different moral virtues, different psychological components (e.g., thoughts, feelings, motivations, behaviors, and outcomes), acts of commission vs. omission, and conduct within different relational roles.

The extent to which commensurability is an issue depends on your moral theory. Overall, the idea that there is a common feature that makes virtues as disparate as kindness, honesty, and loyalty all instances of “morality” is highly contested (for a simplified overview of proposals, see Table S1; for a review, see Sinnott-Armstrong & Wheatley, 2012). Some researchers reject the idea that there is a “one-size-fits-all” ideal of moral excellence, instead arguing and finding

evidence for the existence of distinct varieties of moral personality (e.g., brave, caring, or deliberative; Walker et al., 2010; Walker & Frimer, 2007; see also Flanagan, 1991). If there is no single “essence” that unifies morality, then there is arguably no common metric upon which moral and immoral acts can be compared. Classical utilitarian consequentialists would argue that moral acts are commensurable based on the resulting balance of pleasure vs. pain, so that every moral and immoral act can be evaluated in a common metric based on its hedonic consequences (see also Schein & Gray, 2018). However, because the consequences of an action may involve complex and potentially infinite chains of events and unintended consequences, this might be impossible in practice (Dahl, 2023; Lenman, 2000) or even in principle (Schwitzgebel, 2024). For example, in *The Good Place* (Season 3, Episode 10), a man gave his grandmother a dozen roses in 2009. This ostensibly kind act lost him four moral points because the roses were grown with toxic pesticides, picked by exploited workers, and created profits for a racist billionaire CEO who sexually harasses his female employees. Thus, even on the seemingly most commensurable of moral theories, accurately comparing the moral worth of two actions might be practically infeasible.

Again, we favor a moderate view. Plausibly, in many cases there is no single best weighting, but approximate judgments remain possible. As an analogy, health and money are often treated as incommensurable. But even if health and money can’t be precisely weighed against each other, extreme cases permit straightforward decisions. Most of us would gladly accept a scratch on a finger for the sake of a million dollars and would gladly pay \$10 to avoid stage IV cancer. Similarly, Stalin was clearly morally worse than Martin Luther King, Jr., even if Stalin had some virtues and King some vices. The unjust firing of a dedicated employee is worse than fibbing to your spouse to get out of washing the dishes.

If correct, moderate incommensurability limits the precision of any possible moralometer, whether flexible or inflexible, and to some extent regardless of moral theory, at least in practice. Vices and virtues (and their facets), and rights and wrongs of different types (and subtypes), will be amenable to only rough comparison, not precise determination in a single common coin.

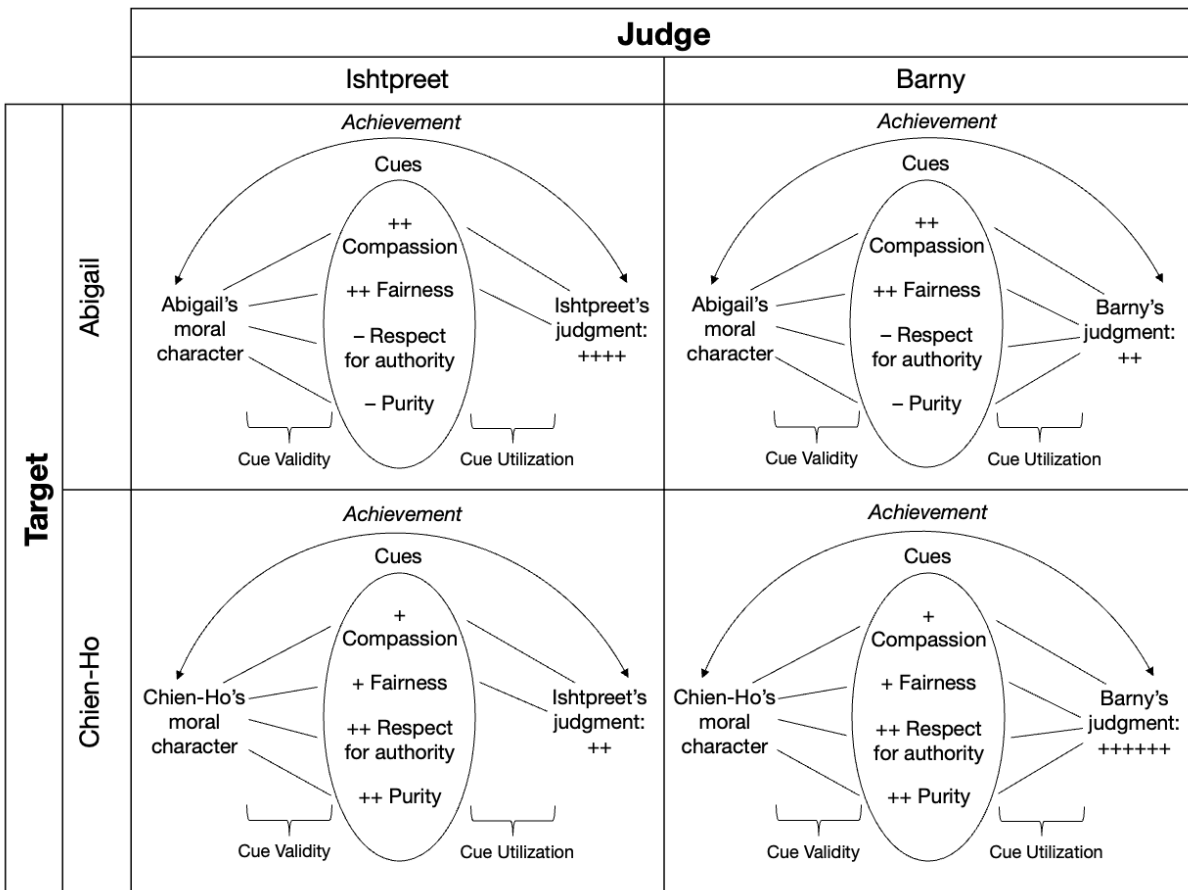
Does the Measure Apply Clearly and Consistently Across People, Groups, and Time?

A measure can only be used to compare people and groups if it retains the same, clear meaning regardless of who is conducting the moral evaluation, and if it applies consistently across people, groups, and time. If a measure produces scores that vary widely and unsystematically depending on either the judge or the target of evaluation, then it is not capturing determinate moral facts about the target. If a measure yields consistent results but those results are difficult to interpret, it likewise fails to constitute a useful moralometer. Flexible measures raise serious challenges concerning both *transparency* (clarity about what is being measured) and *equivalence* (the same features are being measured regardless of judge and target). Fixed measures mitigate these challenges somewhat, but at the potential cost of their *moral relevance*.

Inconsistent Standards. Flexible measures allow judges to use their own moral standards to evaluate a target's general morality. However, "morality" is an extremely broad trait term that can refer to many different behaviors (Hampson et al., 1987). Therefore, there is a substantial risk that different judges' moral ratings will be based on idiosyncratic considerations (Hayes & Dunning, 1997), to the extent that they lack shared meaning systems (Kenny, 1991). Consider, for example, two hypothetical targets (Abigail and Chien-Ho) and two judges (Ishtpreet and Barny; see Figure 1). Abigail is more compassionate and fair but less respectful of authority and pure than Chien-Ho. Ishtpreet only cares about compassion and fairness, whereas Barny considers all four domains to be equally morally relevant. Ishtpreet is therefore likely to

rate Abigail as being more moral than Chien-Ho, whereas Barny will reach the opposite conclusion. If we apply the simplifying (but of course, disputable) assumption that all four virtues are equally relevant to morality, Barny will be correct because he utilizes all four valid cues, whereas Ishtpreet will be incorrect because she only utilizes two out of the four valid cues (Brunswik, 1956; Funder, 1995). Even in less extreme cases where judges agree on which traits are morally relevant, they might not weight the traits similarly.

Figure 1
Moral Disagreement Results in the Use of Different Standards for Judging General Morality



Note. Application of Brunswik's (1956) lens model (see the Realistic Accuracy Model, Funder, 1995, for an extension). For this hypothetical example, we apply the simplifying (but of course, disputable) assumptions that the four virtues (1) are in fact equally relevant to morality (that is, they are all valid cues) and (2) are commensurable, such that each + counts as one positive moral point, and each - counts as one negative moral point. We also assume that judges either use the cues (as indicated by a line from the cue to judgment) or do not use the cues (as indicated by the absence of such line), even though in reality, judges may use different cues to varying extents. Here, Barny validly utilizes all four cues, whereas Ishtpreet only utilizes two out of the four valid cues. Therefore, Barny achieves higher accuracy than Ishtpreet.

Lack of Transparency. When there are no specific descriptive referents, it is unclear which virtues, thoughts, feelings, behaviors, and contextual features are the inputs for a judge's flexible "black box" evaluation of a target's moral goodness. It is therefore difficult to ascertain whether judges are using appropriate moral standards. Such flexible evaluations also provide no concrete information about how we should expect a "moral" or "immoral" person to think, feel, and act in general or in particular situations (Allport & Odbert, 1936; Cattell, 1943; Norman, 1967). Consequently, the resulting measure is difficult to interpret.

We could ask the judges to provide open-ended justifications of their general flexible moral ratings, or to additionally report on various traits that may have informed their general moral judgments (e.g., Schwitzgebel & Rust, 2009; Sun et al., 2025). Such an approach could help reveal some of the considerations that may have informed these flexible judgments. However, this would provide only a rough approximation of which *kinds* of moral considerations were involved, and might not reveal how these various considerations were weighed against each other or crucial morally relevant details of the target and their situation. Moreover, judges may not be fully aware of all the considerations and biases (e.g., stereotypes; Kenny, 2004) that influenced their general moral evaluations.

Costs to Moral Relevance. Fixed measures offer greater transparency about which specific thoughts, feelings, and behaviors are being measured because the criteria and their relative weights are predefined and applied equally to all targets. But beneath the seeming consistency of fixed measures can lie substantial variability in their actual or psychological (i.e., perceived) moral relevance.

Actual Moral Relevance. Costs to actual moral relevance apply if the moral facts are truly different depending on the group and context (as per moderate particularism and

relativism). For example, a fixed moralometer that is based on a specific weighting of a set of moral virtues for which there is historical and cross-cultural consensus (e.g., compassion, honesty, and fairness) could be used to compare people from two different groups on the average of these virtues. But if the traits carry different moral weights in those different groups, or if some group-specific virtues are not included in that composite (e.g., sexual purity, religiosity), the same fixed moralometer would have a different moral meaning when applied to people from the two groups. Similarly, the moral goodness of donating \$10 to charity might differ depending on whether the participant is wealthy, middle class, or impoverished. Therefore, even if a fixed measure captures the same psychological referents at some level of description, there is no guarantee that such a measure can accurately reflect the moral goodness of people in different groups, situations, or across time.

Psychological Moral Relevance. Independently of whether a fixed measure employs an objectively correct conception of morality, if this conception deviates too much from a target's personal understanding of morality, the measure might also lack *psychological moral relevance* (Meindl & Graham, 2014). That is, fixed measures risk dissociation from morality-related psychological processes and consequences such as blame, guilt, and admiration. This concern arises especially when fixed measures are derived from philosophical frameworks or from a culturally- or historically-specific commonsense morality and applied to targets in a different context.

For example, suppose that, after much philosophizing, researchers settle on a utilitarian solution, defining morality as acting in ways that impartially maximize welfare to all beings in the present and future (Kahane et al., 2018). Accordingly, it is morally better to donate \$20,000 to help the global poor than to spend that money on your child's college tuition. A person who is

extremely kind to their loved ones but who never donates to charity is morally worse than someone who is a jerk to the people around them but saves two lives per year by donating 10% of their income to highly-effective charities. It is even selfish to keep both kidneys when only one is needed for survival and the other kidney could save a stranger's life. Such conclusions might be philosophically well-justified; however, if they represent a substantial departure from ordinary moral intuitions, they will be foreign and unpalatable to the average person. If so, then the measure will fail to capture the psychology of morality as people themselves understand it.

For some purposes, it might not matter that a morality measure lacks psychological moral relevance. However, much of what is interesting about human morality depends on its psychological meaning. For example, whereas people who are moral in the sense of embodying widely-accepted character traits such as compassion, honesty, and fairness tend to be better-liked and respected (Goodwin et al., 2014; Hartley et al., 2016), there appear to be social costs of making utilitarian decisions (Everett et al., 2018; Law et al., 2022). A moralometer that diverges too much from everyday conceptions of morality might cease to be practically useful. Flexible measures have an advantage here, but if judges disagree on what constitutes morality, one judge's flexible judgment may not be psychologically relevant to other judges.

Summary

In sum, an intimidating array of conceptual challenges emerge in the quest to measure general morality. Researchers who wish to claim that they are measuring people's overall moral goodness or badness must defend the assumptions that there are general moral facts and that the measure accurately identifies and weighs these facts. For the reasons described, any purely flexible measure will be untenable. Fixed measures are better, but daunting challenges remain. Given the immense philosophical disagreement about such matters, we regard it as unlikely that

any proposed moralometer that claims to measure a person's objective levels of moral goodness would pass conceptual muster in the view of most theorists. Moreover, in the unlikely situation that all of these requirements were fulfilled, the resulting measure may not be recognizable or accepted as being morally relevant to all people who are being measured. As discussed in Supplemental Material, Section 3, similar concerns arise for more specific moral traits, such as compassion and honesty, though arguably to a lesser degree.

Methodological Requirements

Even if we were able to overcome these conceptual difficulties, additional challenges arise for the practical development of a measure that is sufficiently sensitive to variations in general morality (Briggs et al., forthcoming). We see four main classes of methods for constructing a moralometer: self-report, informant report, behavior, and biological markers. Each of these methods raise issues of bias, diagnosticity, and feasibility (see Table 2). Although each of these challenges apply to some extent to the measurement of most psychological traits, we argue that they pose particular challenges for the measurement of general morality (an issue we return to in the Discussion).

Four Approaches to Morality Measurement

Self-Report

Self-report is the bread and butter of personality assessment (Soto & John, 2017; Soto, 2019). To find out how moral a person is, perhaps we could just ask them. For example, the Moral Characteristics Questionnaire (Furr et al., 2022) contains self-report measures of global moral character, compassion, fairness, honesty, loyalty, respect, and purity.

Table 2

Methodological Requirements for Measuring General Morality using Self-Report, Reputation-Based, Behavioral, or Biological Measures

Methodological requirement	Self-report	Reputation	Behavior	Biology
Judges have full information about the target's relevant characteristics.	✓	!!	-	-
Judges correctly use this information to form an unbiased impression of the target's traits.	!!	!	-	-
Judges are willing to truthfully report their impressions of the target.	!!	!!	-	-
The measure represents a holistic evaluation of the target's general tendencies.	✓	✓	!	!!
The measure is feasible enough to be implemented at scale.	✓	!	!!	!!

Note. - = Not applicable; ✓ = Requirement is likely satisfied; ! = Significant difficulty; !! = Major difficulty. Reputation = assume a best-case scenario where there are multiple judges who know the target from different domains of life. See Table S3 in Supplemental Material, Section 3 for an extension of this framework to compassion and honesty.

Informant Report

Friends, family members, romantic partners, acquaintances, coworkers, and teachers have many opportunities to observe our behavior and form moral impressions of us. Most self-report questionnaires can be adapted to the informant perspective, simply by substituting the target's name (e.g., "Geoff tends to act morally", 1 = *strongly disagree*, 5 = *strongly disagree*); Sun & Goodwin, 2020). Therefore, instead of asking people to report on their own moral goodness, we might be able to rely on other people's impressions.

Behavioral Measures

Instead of relying on human judges' subjective impressions, we could directly observe what people do. Behaviors can be assessed to some extent via self-report or informant report (e.g., Rushton et al., 1981). For the purposes of this discussion, however, we emphasize direct observation of "actual behavior" (Baumeister et al., 2007), which we define, in line with Furr

(2009), as explicit “verbal utterances (excluding verbal reports in psychological assessment contexts) or movements that are potentially available to careful observers using normal sensory processes” (p. 372). Implicit and cognitive-attentional measures such as eye movements or reaction times in the Implicit Association Test also count as behavioral, but tend to correlate weakly and inconsistently with real-world moral behaviors (e.g., Kurdi et al., 2019).

Actual behavior can be measured via (a) lab-based or online behavioral tasks (e.g., the fairness of a person’s dictator game allocations; Zhao et al., 2017; cheating tasks; Gerlach & Teodorescu, 2022; children’s helping, comforting and sharing behaviors; Davidov et al., 2013; Warneken & Tomasello, 2006; Ziv & Sommerville, 2017), (b) wearable devices (e.g., kindness-related behaviors captured via audio recordings of everyday conversations; Bollich et al., 2016), (c) digital footprints (e.g., hate speech expressed in social media forums; Kennedy et al., 2022), and (d) objective personal records (e.g., transaction records to measure charitable giving or meat-eating; Ebert et al., 2021; Schwitzgebel et al., 2023; arrest and conviction records to measure criminal behavior; Dodge et al., 2015). Whereas some behavioral observation methods still rely on human judgment (e.g., when judging whether a child “comforted” an adult in the lab or whether a person “expressed gratitude” in an audio recording), other behaviors can be objectively captured (e.g., via transaction records).

Biological Measures

Finally, given that all psychological traits have a biological basis (DeYoung et al., 2022; Krueger & Johnson, 2021), perhaps we could develop a moralometer based on genetic markers, brain structure and function, or physiological measures (e.g., skin conductance, hormone levels).

Can Human Judges Be Trusted?

The validity of self- and informant reports of general morality depends on the extent to which a person is *able* and *willing* to accurately report on their own or others' morality (respectively).

Can Judges Form Accurate Moral Impressions?

Two general factors influence people's ability to form accurate moral impressions of themselves and of others: the *availability* and *utilization* of valid information about relevant thoughts, feelings, motivations, and behaviors (for a more detailed model, see Funder, 1995). The broader the trait, the more difficult it is to meet these requirements. For example, assuming that morality comprises multiple traits (e.g., compassion, honesty, and fairness), one must have access to information about and be able to form accurate impressions of each of the constituent traits to correctly assess a person's general morality.

Information Availability. On the ability side, people are in a privileged position of having continuous and relatively direct access to not only their behaviors, but also their thoughts, feelings, motivations, and relevant context. This means that—barring memory lapses (Carlson et al., 2020) or perceptual deficiencies (Wright et al., 2020)—people are less likely than their acquaintances to lack relevant information that could be used to judge their own overall moral goodness. In contrast, informants are not always with the target and generally have less access to the target's relevant thoughts, feelings, and motivations. Moreover, because people play different relational roles in our lives, our romantic partners, coworkers, friends, and teachers will each observe us in limited and possibly unrepresentative contexts. Thus, each informant report is based on only a fraction of the information that is available to the self. Such self–other differences are likely exacerbated for certain moral traits (Thielmann et al., 2017). For example, it is arguably more difficult for a person to hide how (un)compassionate they are than how

(dis)honest they are, because judging someone's honesty requires knowing both (a) the truth and (b) the fact that someone is misrepresenting it.

Information Utilization. Even if all judges had access to all information that would be relevant to judging a target's general morality, people can form different opinions based on the same information, for example if their attention or inferences are biased (as detailed by various models of person perception and personality judgment; Funder, 1995; Kenny, 1991, 2004). Despite the self's informational advantages, self-serving ego-protective biases (Paulhus & John, 1998) may prevent people from being able to admit to themselves that they have substantial moral shortcomings. According to the Self-Other Knowledge Asymmetry Model (Vazire, 2010), others have an advantage compared to the self when judging highly evaluative traits if the traits are also highly observable. The logic is that other-perceptions are less likely to be distorted by ego-protective biases because there is less at stake when judging others' overall moral goodness (vs. our own). Also, more information is not always better; for example, excessive focus on internal states (e.g., good intentions) might bias self-evaluations, whereas informants relying on observable moral behaviors might form evaluations founded more on concrete evidence (Klein & Epley, 2017).

However, informants' ability to form impartial judgments can depend on their relationship with the target. Being well-acquainted with targets is associated with greater accuracy and with less positivity bias, whereas liking them is associated with lower accuracy and greater positivity bias (Wessels et al., 2020). Thus, the ideal informant is someone who knows you well, but does not particularly like you. However, there is generally a tradeoff between familiarity and liking, such that those who are better able to provide an impartial judgment also have less information (Wessels et al., 2020). And, in practice, people tend to select informants

who have positive perceptions of them (Leising et al., 2010). Informant reports may therefore be contaminated not by self-serving biases, but by “pal-serving biases” (Leising et al., 2010). In fact, informant reports tend to be even more positive than self-reports (Kim et al., 2019).

Given these differences in bias and information, if informant reports are used, at a minimum, it is important to hold the informant type constant between targets (e.g., using only friends as informants). Given that any single informant has an incomplete and potentially idiosyncratic perspective on a target’s moral behaviors (Pringle et al., 2024; Smith & Apicella, 2020), however, sampling multiple informants across different relational roles (e.g., recruiting one friend, one romantic partner, and one coworker to report on each target) would be ideal. Indeed, a structural advantage of informant reports over self-reports is that it may be possible to gain more reliable measurements by aggregating across multiple informants (Hofstee, 1994). Thus, even if the self and others were equally accurate, an aggregate of multiple informant reports—which reflects the target’s moral *reputation*—could potentially result in a more accurate moral reading than a single self-report.

Rating Inversions. If everyone inflated their self- or informant reports to a similar extent, we could at least draw conclusions about relative morality. However, self- and other-evaluation biases could operate differently for different informants and targets. This would at least add noise and might even sometimes produce rating inversions in which more moral people judge themselves as less moral than do truly less moral people. Lending plausibility to this idea, Sedikides and colleagues (2014) found that prisoners rated themselves more favorably than the average community member on all prosocial characteristics that were assessed (e.g., *moral, kind to others, honest, trustworthy, dependable*), except for law-abidingness.

Why might rating inversions occur? First, we speculate that accurate judgment of one's own moral goodness might to some extent depend on having good moral character. For example, moral people might be more willing to non-defensively recognize their own moral weaknesses (Smith & Kouchaki, 2018), whereas less moral people might arrogantly assume that they're in the right. Similarly, there could be a moral Dunning–Kruger effect (Kruger & Dunning, 1999), such that the least moral people lack the moral knowledge necessary to understand that they deserve low ratings. For example, someone with little knowledge of the dynamics of sexism might regard themselves as entirely non-sexist, while someone more familiar with those dynamics, and who as a result is objectively less sexist, might see traces of sexism in themselves. Second, if moral people are less able to live up to their lofty moral standards, whereas less moral people aim for and achieve moral mediocrity (Schwitzgebel, 2019), moral people might have harsher self-assessments than do less moral people. Finally, some moral traits—perhaps especially humility—might be inconsistent with high degrees of self-praise (Robinson, 2020).

Similar concerns arise for informant reports. If more moral targets are more willing to transparently admit their moral failings, the informants of more moral targets might have more knowledge of moral targets' less observable moral flaws (whereas less moral targets might successfully hide their immoral behaviors and thoughts). Based on evidence of personality similarity among couples and friends (Youyou et al., 2017), people with low moral standards might also tend to have informants with low moral standards. If morally ignorant people tend to have similarly morally ignorant informants, then we might also see an interpersonal moral Dunning-Kruger effect (e.g., a Nazi whose informants are other Nazis).

Such issues are not unique to the measurement of general morality. For example, reference group effects may explain why Mexicans self-report being less sociable than

Americans, even though objective measures based on audio recordings of everyday life (e.g., time spent talking to others) show the opposite (Ramírez-Esparza et al., 2009). However, as detailed above, we suspect that there are several additional mechanisms (e.g., differences in moral ignorance, moral humility, and transparency) that might contribute to a particularly high risk of rating inversions in self- and informant-reports of overall moral goodness.

Are Judges Willing to Accurately Report their Impressions?

Even if people did have perfect knowledge of their own general morality, convincing them to candidly report their judgments is another matter—especially in non-research contexts where one’s responses might have real consequences (e.g., a job application or dating app questionnaire). For example, Anglim and colleagues (2017) found that those who completed the HEXACO in the context of a job application (vs. a low-stakes research context) reported much higher levels of sincerity ($d = 0.50$) and fairness ($d = 0.86$). The effects of socially desirable responding need not be unidirectional: To say that you are extremely immoral seems obviously socially undesirable, but describing yourself as “extremely moral” could seem immodest or dishonest (Choshen-Hillel et al., 2020). Social desirability biases may also operate in different ways for different traits and contexts (e.g., someone who is applying to be a salesperson might strategically report that she is only 70% honest).

It is not clear that informants would be any more willing to candidly report their moral evaluations of others—especially if they have a negative perception of the target. Even if informants are assured of confidentiality, they may still worry that the targets would find out what they said. Informants may also feel uncomfortable about “telling on” the target, especially in consequential contexts. Even independent of consequences, the mere act of casting moral judgment on others might feel self-righteous or “mean” (Sun et al., 2022; Yudkin et al., 2023).

Such concerns seem less likely to apply to informant reports of morally neutral traits, such as extraversion. Together, these issues also aggravate concerns about sampling bias (such that informants who have more negative perceptions of targets would be less likely to respond) and the ability to recruit even one—let alone several—willing informants.

Could Behavioral or Biological Measures Do Better than Human Judges?

Given the issues of relying on human judgment, could direct behavioral observation or biological measures be the answer? Behavioral measures are appealing because they seem to more directly and objectively reflect a person's moral conduct. This directness also means that there is less of an inferential gap between the measure and the behavior. Whereas “strongly agreeing” that a person is “moral” or “helpful and unselfish with others” could be interpreted in many ways, transaction records that show that person has spent \$2,000 on verifiable charitable donations that year (e.g., Ebert et al., 2021) are concrete and specific. Similarly, biological measures seem to produce an objective measurement (e.g., a genetic profile, the volume of a specific brain region, or skin conductance) that generally cannot be hidden, faked, or distorted by human perception. Despite these apparent advantages, behavioral and biological measures have major disadvantages when it comes to representativeness and feasibility.

Does the Measure Represent a Holistic Evaluation of the Target's General Moral Tendencies?

Although describing a person as being “helpful and unselfish with others” is in some respects vague, such a broad question targets the full range of relevant characteristics, as well as the relevant motivations and contextual features. Self- and informant reports of broad moral character traits allow judges to holistically and efficiently summarize across large amounts of relevant information about what a person is generally like across multiple situations. In contrast,

behavioral and biological measures present vexing inferential gaps concerning their relevance to a person's overall moral goodness.

Non-Moral Reasons for (Im)Moral Behavior. First, people may engage or fail to engage in ostensibly moral behaviors for non-moral reasons. If a virtue ethics framework is correct, then any assessment that focuses only on behavior (without considering the accompanying thoughts, feelings, motives, and situational constraints) will fail to validly measure a person's moral character (Fowers et al., 2021; Wright et al., 2020). For example, whereas someone who primarily follows the rules to avoid punishment would cheat when they think they could get away with it, a person who is motivated by considerations of fairness would be more reliably honest. That is, honest behavior can be distinguished from virtuous motivation, and an honest *person* not only acts honestly but would do so in a wide range of conditions and for the right reasons (Miller, 2017; Roberts & West, 2020; Wilson, 2018). Self-report measures can incorporate such motivational requirements in a way that behavioral measures generally cannot (outside the confines of lab experiments that manipulate situational features; e.g., Batson et al., 1988; Zhao et al., 2017). For example, the HEXACO fairness measure (Lee & Ashton, 2018a) asks people if they would be willing to steal a million dollars *if they knew they could never get caught*, and the Truthful Communication Scale (Furr et al., 2021) includes references to being willing to tell the truth *even if it might hurt someone's feelings*.

Considerations of motivation, intention, and context might also have implications for whether a given behavior reflects a general moral disposition (vs. an exception). For example, many vegetarians are primarily motivated by personal health (rather than animal welfare or the environment; Hopwood et al., 2020). If so, vegetarianism might be a weak indicator of general moral behavior. Someone who donates large sums of money to gain social status might not act

generously without social rewards. Some kinds of “(im)moral” behaviors may also reflect environmental context more than a person’s (im)moral disposition. For example, some of the behaviors featured in a widely-used “altruism” scale (Rushton et al., 1981) are relevant only to specific contexts (e.g., “I have helped push a stranger’s car out of the snow” only applies in snowy climates) or are subject to confounds (e.g., “I have given a stranger a lift in my car” may be confounded with neighborhood safety). Similarly, social scientists commonly use donations in the context of economic games as an indicator of generosity (vs. selfishness). However, Carlson and Crockett (2024) question the assumption that keeping money (instead of donating) in such games necessarily reflects selfishness, finding that 47% of online workers report financial need (e.g., being unemployed, struggling to pay medical bills, being unhoused) as the dominant motive for keeping (instead of donating) a bonus payment.

Unrepresentativeness of Single Behaviors. Second, single behaviors observed in the lab (e.g., cheating in a laboratory game) or in everyday life (e.g., donating, meat-eating) are unlikely to be *representative* of a person’s general morality, or even their possession of a specific moral virtue. A person who doesn’t engage in the particular moral acts measured in a study could be more moral overall than other people who engaged in the measured behaviors. Arguably, however, if a reasonable number and range of behaviors that pertain to a specific moral trait (e.g., compassion) are measured, it might be more parsimonious to conclude that these behaviors reflect an underlying compassionate disposition that would also cause the person to enact compassionate behaviors that were not specifically measured (in line with a reflective latent variable model; Edwards & Bagozzi, 2000). Per the “duck test” (Block, 1993; Funder, 1995), something that looks, walks, and quacks like a duck may not be a duck, but it probably is. The burden of proof eventually falls on the skeptic to show otherwise. However, the duck test is

likely to be much harder to satisfy for general morality than for a specific moral trait because of the wide diversity of dimensions of morality and the likelihood that the dimensions of morality are less well correlated than facets of a narrower trait such as compassion. Therefore, even if several compassion-related behaviors were measured, a measure that lacks honesty- or fairness-related behaviors (for example) would fail to distinguish someone who is compassionate but dishonest and unfair from someone equally compassionate but much more honest and fair.

Tradeoffs Between Feasibility and Diagnosticity. There may also be a tradeoff such that the behaviors that are easiest to observe (*high feasibility*) are less clearly indicative of a person's overall moral goodness (*low diagnosticity*). It is straightforward to observe whether a person misreports the outcome of a coin flip to earn more money in a laboratory study (Pascual-Ezama et al., 2020), but it's disputable whether observations based on such artificial behavioral paradigms—behaviors that are typically targeted at the experimenter or hypothetical “raceless, genderless strangers” (Hester & Gray, 2020; see also Earp et al., 2021; Schein, 2020)—reveal anything about a person's general moral conduct in the real world (cf. Dai et al., 2018; Kröll & Rustagi, 2016; Schild et al., 2021). Conversely, some of the most morally consequential actions—donating a kidney, heroically saving a life, or killing someone—might be highly diagnostic but unlikely to occur at high enough base rates to be useful for distinguishing among the 90+% of morally typical people. Barring extensive investigations and audits, more common moral violations such as cheating on a spouse, sexual harassment, or underpaying taxes by thousands of dollars would also be difficult to directly measure, given that people typically want to hide such immoral behaviors. Indeed, even attempting to measure such behaviors might constitute an unethical invasion of privacy.

One could also attempt to measure more subtle types of moral behavior. For example, the Electronically Activated Recorder (EAR) records brief audio snippets of people's lives at random or pre-programmed intervals (e.g., 30 s every 10 minutes) for one or two weeks (Mehl et al., 2001). Human coders can then code these recordings for specific acts that researchers have pre-specified as being morally relevant (e.g., showing sympathy, offering help, criticizing, being condescending; Bollich et al., 2016). Hours of audio recordings provide a good sense of how people generally treat the people around them. However, given that there is still room for interpretation, even teams of three or four coders per person often achieve only modest levels of inter-judge reliability (Bollich et al., 2016; Sun, 2020). In addition, although the EAR is well-suited for capturing traits that are primarily defined by audible behaviors in social interactions (Mehl, 2017; Sun et al., 2020), such as compassion, it is ill-suited for capturing traits that cannot easily be heard (e.g., dishonesty).

The Inadequacy of Existing Biological Measures. Compared to direct behavioral observation, it might be tempting to think that some biological measures (e.g., genetic markers, brain structure or function, physiological measures) could be more reflective of broad underlying causal mechanisms that are responsible for a range of moral behaviors—a person's moral “essence” or “potential”—what they are truly capable of, beyond the surface-level behaviors that have only been observed in a limited number of situations. However, all existing biological measures are—and for the foreseeable future will remain—either too indirect or too noisy to ground meaningful conclusions about general individual-level morality (for more discussion, see the Supplemental Material, Section 2).

Is the Measure Feasible Enough to be Implemented at Scale?

Even if it were possible to develop an ideal moralometer, participant burden, unwillingness, privacy, or cost may pose practical barriers to the feasibility of implementing such a measure. Self- and informant-report questionnaires can efficiently summarize large amounts of information about a person's moral goodness. Self-report is the most feasible option, as this method only requires one judge per target. In contrast, informant methods ideally require multiple judges to mitigate ignorance and idiosyncratic variation, recruitment of which can be difficult. When sampling multiple informants, it might also be difficult to ensure that each target is rated by same set of informant types (e.g., one friend, one romantic partner, and one sibling), as not everyone has certain relationships. Practical constraints thus limit the feasibility of an ideal moral reputation-based measure.

Behavioral measures are even less feasible. Each of the main methods for behavioral observation require a high level of compliance from the participant and could be perceived as intrusive. Behavioral coding methods (e.g., Bollich et al., 2016) are prohibitively labor-intensive to implement at scale, although advances in artificial intelligence (e.g., automated extraction of behavioral cues from videos; Barto et al., 2017; Mast et al., 2015) may help in the future. Given the difficulty of directly measuring even one behavior, it would be extremely difficult to implement an ideal behavioral moralometer that comprises a composite of several instances of multiple diagnostic behaviors. Such feasibility issues are essentially insurmountable if the goal is to accurately assess an individual's general morality, which would require observing many behaviors diagnostic of multiple aspects of a person's (im)morality.

Biological measures are also infeasible. Some are prohibitively expensive (e.g., MRI, cortisol assays) or inconvenient (e.g., EEG, galvanic skin response) to implement at scale. People

might also be understandably averse to providing their genetic or neural data for the purposes of moral character assessment.

Summary

In sum, even setting aside broad conceptual issues of the sort discussed in the first half of this article, substantial methodological challenges plague each of these four approaches to measuring a person's general morality. Self- and informant-report measures have advantages of representativeness and feasibility, but rely on judges' ability and willingness to accurately report a target's moral goodness. Behavioral and biological measures are more concrete and "objective," but it is difficult to justify general conclusions about a person's overall moral goodness from a small number of behaviors or biological markers. In combination, these issues reduce the accuracy with which a person's general morality can be measured, to a much greater extent compared to most other psychological traits (see Discussion). Nevertheless, as we discuss below, these methods may still be useful for studying morality-related phenomena (see Discussion).

Discussion

How General Are These Conceptual and Methodological Concerns?

Versions of our conceptual and methodological concerns can be raised for most psychological traits. Thus, it might seem that we have made a case for blanket skepticism about psychological trait measurement. However, we hold that the challenges are less severe for most other traits than for general morality. To see how, we consider the general lessons for measuring narrower aspects of moral functioning and non-moral traits.

Lessons for Measuring Narrower Moral Psychological Phenomena

So far, we have painted a pessimistic picture of the conceptual and methodological challenges of measuring a person's overall moral goodness. But, despite the daunting challenges, we are not complete pessimists about the measurement of moral psychological phenomena. Therefore, in Table 3, we outline several more tractable alternatives to measuring general morality. For each of these alternatives, some of the conceptual and methodological challenges in measuring general morality can either be circumvented to a significant extent or no longer apply, though in different degrees and in different respects for the different approaches.

First, although we have argued that ambitious, general claims about a person's overall moral goodness are unwarranted, it may be reasonable to draw conclusions that are circumscribed to specific operationalizations of moral goodness. Conceptually, measuring "general morality" or "overall moral goodness" requires choosing the correct ethical framework. No such requirement applies if researchers specify that they are measuring "deontological rule adherence, operationalized as the fulfilment of five specific duties" or "commonsense morality, operationalized by an equally-weighted set of moral virtues that people typically consider to be morally relevant in North America in 2023," and calibrate their conclusions accordingly. Even if the link between commonsense morality (for example) and objective moral truth turns out to be tenuous, we can still learn something about the psychological causes and consequences of commonsense morality in a particular cultural-historic context. For example, those who are seen as being more moral based on a composite of common moral virtues (e.g., compassion, honesty, fairness) tend to report being happier (Sun et al., 2025).

Table 3*Alternatives to General Morality*

Description	Example(s)
Specific operationalizations of (im)morality Morality in line with a specific philosophical framework Commonsense morality , as defined by what is consensually considered to be morally relevant within a particular population	Deontological rule adherence, operationalized as the fulfilment of five specific duties An equally-weighted composite of benevolence, respectfulness, general morality, dependability, loyalty, honesty, interpersonal fairness, and fraud avoidance (Sun et al., 2023)
Specific manifestations of (im)morality Specific virtues/vices Specific behaviors Moral traits or behaviors within a specific context/role	Compassion, honesty, moral courage, fairness, loyalty, “dark” traits (e.g., psychopathy, sadism, Machiavellianism; Moshagen et al., 2018), honesty-humility (Lee & Ashton, 2018b), guilt-proneness (Cohen et al., 2012) Charitable donations, expressing sympathy, cheating on one’s partner Wisdom, courage, humanity, justice, temperance, and transcendence expressed while parenting vs. working (Bleidorn & Denissen, 2015)
Moral psychological components that do not necessarily reflect actual (im)morality Subjective moral perceptions of one’s own or others’ general morality based on flexible measures that do not specify moral content Morally-motivated behavior based on moral convictions (irrespective of whether it is in fact a morally good behavior) Moral reasoning: How people explain their moral judgments Moral values: The principles and convictions that guide moral judgment and conduct Moral (in)congruence in aligning one’s behaviors with one’s own moral values (irrespective of whether these values are in fact morally good)	The General Morality subscale of the MCQ (Furr et al., 2022) Moral vegetarianism (Hopwood et al., 2020), pro-life/pro-choice activism (Vanderford, 1989), violent opposition to compromise over issues considered sacred (Ginges et al., 2007) Semi-structured moral judgment interview (Colby & Kohlberg, 1987); Defining Issues Test (Rest et al., 1997) Moral Foundations Questionnaire – 2 (Atari et al., 2023); Moral Expansiveness Scale (Crimston et al., 2016); Two-Dimensional Utilitarianism Scale (Kahane et al., 2018) Using pornography or eating meat despite believing that doing so is immoral (Grubbs et al., 2019; Schwitzgebel & Rust, 2014)

Second, it is more feasible to meet the requirements discussed above (see Tables 1–2) when the goal is to measure more specific manifestations of moral goodness. For example, ratings of trait breadth show that compassion and honesty are narrower concepts than morality

(Hampson et al., 1987). Accordingly, there is relatively less room for reasonable conceptual disagreement about compassion and honesty than about general morality, and the measurement targets are specific. If so, the conceptual and methodological challenges of measuring such specific traits could be somewhat less severe (though substantial challenges still exist; see parallel discussion in Supplemental Material, Section 3, and Tables S2–S3).

Challenges of weighting can be further reduced if the goal is to measure moral conduct within specific contexts (e.g., in the workplace vs. at home), or specific moral behaviors (e.g., charitable donations), without attempting to generalize to a person's general morality. For example, if the aim is to understand whether undergraduate philosophy instruction can influence students to eat less meat (Schwitzgebel et al., 2023), this single behavior does not need to be diagnostic of the target's general moral tendencies—or even their general tendency to be compassionate. Taking a more specific approach would not only be more tractable, but might also be more useful. For example, in partner selection or hiring, we are probably more interested in whether that person will be compassionate or honest to *me*, or compassionate towards their *coworkers* or *patients* (Law et al., 2022; Lukaszewski & Roney, 2010).

Third, many aspects of moral functioning that do not necessarily reflect objective moral goodness nevertheless have important intrapersonal, interpersonal, and intergroup consequences. Subjective perceptions of others' morality are more important than perceptions of competence or warmth in impression formation (Goodwin et al., 2014; Hartley et al., 2016). Pornography users are more distressed if they believe that using pornography is morally wrong than if they experience no such moral incongruence (Grubbs et al., 2019). Strong moral convictions can drive opposite behaviors (e.g., pro-choice vs. pro-life activism; Vanderford, 1989) and cause violent opposition to compromise over sacred values (Ginges et al., 2007). Such moral psychological

phenomena can be captured using self- and informant-reports that are appropriately described as reflecting moral perceptions, moral (in)congruence, and morally-motivated behavior, respectively, without claiming that they reflect actual moral goodness. And, when the focus is on subjective perceptions, we no longer need to assume that judges have formed a complete and unbiased picture of their own or others' moral goodness. For example, if the aim is to compare the moral self-perceptions of people in two demographic groups, greater self-enhancing bias and ignorance of moral facts in one group might be potential explanations of accurately observed patterns of higher self-opinion in that group, rather than sources of measurement error.

Although researchers applying these alternative approaches should be, and typically are, wary of framing their conclusions as pertaining to a person's overall morality or general moral character, the conceptual and methodological reasons for wariness differ depending on the nature of the approach, as illuminated by our framework.

Do These Concerns Apply to the Measurement of Non-Moral Traits?

To see how these conceptual and methodological challenges arise more severely for most moral traits than for most non-moral traits, we consider two example non-moral traits: extraversion and well-being (see Tables S4–S5). We choose these examples because extraversion is a fairly uncontroversial concept, whereas well-being resembles morality in being broad, value-laden, and contentious.

Extraversion. Extraversion is the tendency to be outgoing, talkative, assertive, and energetic. Conceptual challenges include that the particular manifestations can vary by individual and group (e.g., Olaru et al., 2019), facets might reasonably be given different weights, and lay conceptions can depart from expert understandings (Kaufman, 2014). Overall, however, there is

much less disagreement about the components of extraversion, its measurement, and its coherence as a construct (Wacker & Smillie, 2015) than about general morality.

Methodologically, participants normally have extensive information about their extraverted or introverted tendencies, are able to form fairly accurate impressions of these facts about themselves, and are normally willing to report their impressions. Being a relatively observable and not highly evaluative trait (Vazire, 2010), self-ratings, informant ratings, and behavioral observations tend to correlate relatively well (Connelly & Ones, 2010; Tackman et al., 2020). Self-reports of aspects of extraversion appear to have adequately specific referents (e.g., “Is dominant, acts as a leader”; “Is full of energy”; Soto & John, 2017) and validated personality questionnaires are diagnostic of the target’s general tendencies in everyday life (Fleeson & Gallagher, 2009) and implementable at scale.

Well-Being. Philosophers use the term *well-being* to describe what is intrinsically good for someone and what it means for that person’s life to go well (Chappell & Meissner, 2023). Being a similarly broad, value-laden concept, well-being is perhaps as conceptually controversial as morality. For example, different theories characterize well-being in terms of the balance of positive vs. negative emotions (*hedonic theories*), getting what you want (*desire satisfaction theories*), or the attainment of objectively valuable goods such as the capacity for meaningful relationships, authenticity, and self-knowledge (*objective list theories*; Crisp, 2021; DeYoung & Tiberius, 2023; Margolis et al., 2021; Parfit, 1984). Well-being could also mean somewhat different things in different cultural or religious traditions (Hitokoto & Uchida, 2015; Oishi et al., 1999, 2013; Tiberius, 2004). Nevertheless, self-report measures derived from different theories of well-being tend to converge on similar answers (Disabato et al., 2016; Goodman et al., 2018; Margolis et al., 2021), perhaps because “objective list” goods such as positive social

relationships and meaningful activity tend to be widely desired and produce hedonic benefits (Bishop, 2015; Chappell & Meissner, 2023). If so, this limits the measurement impact of the different conceptualizations.

Still, although scientists now broadly agree that well-being can be measured (while continuing to debate how best to measure it; DeYoung & Tiberius, 2023), skeptics within philosophy remain. Just as it is unlikely that we can come up with an exact recipe for how to weight various components of morality, weighting is also a particular challenge for the measurement of well-being. For example, Scanlon (1998) worries that though we may manage to agree on core components of well-being, it is unlikely that we can agree on how much weight to assign to each. Taking a more extreme view that emphasizes the importance of person-specific weights, Hausman (2015) argues that well-being must be measured by aggregating goods in a person's life in a way that respects individuality. He argues that existing measures fail to do so, and that this makes it largely unrealistic to reliably compare well-being states on the population level. In response, Alexandrova (2017) largely agrees that well-being is not measurable in this highly demanding, expansive, "all-things-considered" sense. Instead, Alexandrova argues that the science of well-being can and should focus on measuring context-specific notions of well-being (i.e., how a person is doing given their circumstances; e.g., children, working mothers, caretakers of the ill, etc.). Similarly, Tiberius and Haybron (2022) argue that different conceptions of well-being will generate measures that correlate differently with different predictors and outcomes.

To the extent well-being depends on internal experience and individual desires, self-report will arguably be a better option than informant reports or behavioral measures. People are probably not as motivated to misrepresent their well-being as their morality. But concerns can

still be raised about the accuracy of self-report. For example, some people—especially in the U.S.—might overreport their happiness (Haybron, 2008). If well-being is operationalized in terms of desire satisfaction, accurate measurement might require that people know what their true desires are, their relative weights, and the extent to which they are satisfying these desires. Self-knowledge of this sort might not be perfect (DeYoung & Tiberius, 2023). People’s well-being judgments are also susceptible to judgment biases such as recent salient events and their current mood (Conner & Barrett, 2012), and do not track the negative effects of deprivation and oppression as much as one might expect (e.g., Biswas-Diener & Diener, 2006; Tang et al., 2019). Despite these issues, researchers appear to be mostly justified in assuming that self-reports of well-being are approximately accurate, in part because of the convergence among different self-report measures of well-being, in part because self-report well-being measures mostly correlate as expected with other psychological and situational variables (Diener et al., 2018), and in part because of “network effects” in which different aspects of well-being tend to reinforce each other (Bishop, 2015).

The Ethics of Measuring a Person’s General Morality

Would it be a good thing for a half-decent moralometer to be widely available, even if the results were only approximate? People could use moralometers to make better decisions about who to trust or shun, and a great deal of harm could arguably be avoided if a moralometer was used to ensure that only highly moral people are elected into positions of power. For example, after reviewing associations between measures of moral character and (un)ethical behavior in the workplace, Cohen and Morse (2014) concluded that “An obvious implication...is that individuals low in moral character should be avoided in organizations and other social settings, lest one be cheated, defrauded, or betrayed by them” (p. 56). The widespread availability of

moralometers might even help improve the population's morality by helping researchers assess which interventions work or by encouraging people to become more moral via the social rewards or punishments associated with the routine use of moralometers in the personal, professional, and civic spheres. But we advocate caution because any tool that is taken seriously as an accurate measure of a person's general morality has substantial potential for misuse. For example, Hare (1998), creator of the Psychopathy Checklist-Revised (PCL-R), has expressed concerns about its potential for misuse in the criminal justice system (e.g., to deny parole to inmates), especially when judgments are made by non-clinicians or "hired guns" who may lack sufficient training, experience, or integrity to score the items in an unbiased manner. At best, narrow judgements ("this person is a high recidivism risk, based on the PCL-R"; Salekin et al., 1996) will likely be both more practically targeted and less problematic than broad judgments about overall moral goodness.

First, given the conceptual challenges described above, some forms of moral disagreement will be reasonable. Anyone whose conception of morality reasonably diverges from the conception employed in the construction of the hypothetical general moralometer would then likely be unjustly misjudged. Dystopias such as *Brave New World* (Huxley, 1932), *1984* (Orwell, 1949), and *The Handmaid's Tale* (Atwood, 1985) vividly illustrate the hazards of government enforcement of misguided moral schemes. Implementations and prospective implementations of a "social credit system" in China raise similar concerns (Creemers, 2018; Liang et al., 2018). Second, any realistic measure will have substantial inaccuracies. If decision-makers employ the measure in a high-stakes context (e.g., hiring, educational admissions, loans, immigration, dating), inaccurate moralometer readings could cause truly moral people to be unfairly excluded from opportunities. Moreover, per Goodhart's Law—"when a measure

becomes a target, it ceases to be a good measure” (Strathern, 1997)—unless the moralometer is based wholly in something that cannot be faked, people will find ways to exploit its inaccuracies. Consider, for example, the reputation consultants in *Black Mirror: Nosedive* who helped clients derive strategies for enhancing their ratings.

But don't we already always morally evaluate the people around us? Employers judge employees, dating-app users judge first dates, and voters judge politicians. Such judgments are probably highly inaccurate and subject to systematic biases (e.g., racial stereotypes; Welch, 2007). Wouldn't a half-decent moralometer then improve the accuracy of what we already do? To this objection, we reply that a moralometer presented as a scientific tool risks aggravating the problems of misjudgment. First, reasonable people often have appropriate epistemic humility, feeling uncertain about their initial moral judgments and thus being at least somewhat open to new evidence. In contrast, a “science-backed” moralometer risks inappropriately inflating people's confidence and thus their likelihood of acting without seeking further evidence. Second, currently, even if one person unfairly judges you to have a bad moral character, the next person might judge you more positively. This distributes the inaccuracies across the population (though imperfectly, given some systematic biases). The biases of a moralometer, however, will be highly systematic (e.g., if all employers use the same moralometer).

Recommendations for the Responsible Use and Interpretation of Morality Measures

Clearly, some approaches to moral measurement are more (e.g., measuring specific acts of compassion) vs. less (e.g., flexible measures of general morality) tenable. What is important is that researchers exercise thoughtfulness and intellectual humility when developing and communicating the results of moral measurement. Accordingly, we propose five recommendations for the responsible use and interpretation of morality measures (see Table 4).

Table 4*Recommendations for the Responsible Use and Interpretation of Morality Measures*

Recommendation	Rationale
1. Engage with the diversity of moral thought across various philosophical, religious, and cultural traditions.	To encourage appropriate epistemic humility and create better-founded measures of moral goodness.
2. Make value claims and assumptions transparent.	To facilitate critique of those values.
3. Prioritize interpretability.	To allow others to understand and evaluate the measure for themselves.
4. Measure more specific aspects of moral functioning and calibrate conclusions accordingly.	Conceptual and methodological requirements are more attainable or less relevant when the measurement goals are more specific. Calibrating conclusions will help prevent misinterpretation.
5. Consider the stakes before acting on morality measures.	To avoid the ethical issues with moral misjudgment, especially given inevitable measurement error.

1. Engage with the Diversity of Moral Thought

One potential pitfall is that a moralometer based on personal predilections (Meindl & Graham, 2014) might be presented as capturing the “One True Morality” (Dahl, 2023). Even multiple psychologists working together might still have a narrow outlook, given the field’s lack of political and cultural diversity (Arnett, 2008; Duarte et al., 2015; Henrich et al., 2010; Inbar & Lammers, 2012; Redding, 2001). Engaging with the diversity of moral thought across philosophical, religious, and cultural traditions—as well as the complexities and debates within even a single tradition—will likely help psychologists better recognize the assumptions and limitations of their measures, enabling the creation of better-founded measures.

2. Make Value Claims and Assumptions Transparent

Science—and especially moral psychology—is inevitably value-laden (Alexandrova, 2017; Fowers, 2022; Prinzing, 2021). Indeed, an important part of the value and interest of moral psychology research is its capacity to reveal to us why people behave in ways that are in fact

morally good or bad, with the non-value-neutral aim of transforming society for the better.

Instead of attempting to be “value-neutral,” researchers should explicitly acknowledge the values and assumptions built into their operationalization of moral goodness. Being transparent and explicit about those values will facilitate critique of those values and more productive dialogues about the merits of various conceptualizations of morality.

3. Prioritize Interpretability

Researchers should prioritize interpretable measures of morality. If people cannot understand how the moralometer generates its evaluations, it will be difficult to understand what the moralometer is capturing and to trust or challenge its results (Purcell & Bonnefon, 2023; von Eschenbach, 2021). One potential downside is that making the “answer key” available could help people learn how to look good on the test, but this can be mitigated by adhering to our fifth recommendation not to draw conclusions about individuals.

4. Measure Specific Aspects of Moral Functioning and Calibrate Conclusions Accordingly

As discussed above, various conceptual and methodological challenges can either be circumvented to a significant extent or no longer apply when the measurement goals are more modest (see Table 3). Accordingly, we recommend that researchers measure specific operationalizations of moral goodness (e.g., commonsense morality), specific manifestations of moral goodness (e.g., specific acts of compassion), or other aspects of moral functioning that do not necessarily reflect actual moral goodness (e.g., moral self-perceptions, moral incongruence). We encourage researchers to clarify the measurement aim, to assess the extent to which conceptual and measurement challenges (see Tables 1–2) arise given that aim, and to calibrate interpretations and conclusions accordingly by using appropriately narrow, precise labels (see

Table 3). Because otherwise valid measures can be misinterpreted when inappropriately labeled, labels should be chosen carefully to reflect the conclusions that can legitimately be drawn.

5. Consider the Stakes Before Acting on Morality Measures

Even if there are general facts about a person's overall morality, the measurement issues we've discussed are sufficiently severe that correlations between any practically feasible moralometer and a person's actual level of moral goodness are likely to be small to medium sized at best (Doris, 2022; Möttus, 2022). Drawing conclusions about individuals' general morality based on any practically feasible measure is therefore likely to substantially mischaracterize many people's actual morality.

As previously mentioned, however, validity is a matter of degree and what counts as good enough for practical purposes varies with the proposed uses (AERA et al., 2014), the degree of accuracy, and the likely directions and conditions of error. If your goal is to create an entertaining listicle of the relative overall morality of famous historical figures, crude informant ratings would suffice, reliably generating the result that Mahatma Gandhi was overall morally better than Benito Mussolini. In any context with real stakes, however, low accuracy risks unfairly wronging the people who are being evaluated. Accordingly, a measure that is used to select people for employment or college admissions would only be justifiable (if at all) if it is much more accurate, and it would be especially troubling if the measure were, for example, systematically racially biased.

In daily life, we sometimes act on the basis of personal judgments about the overall morality of the people around us. It would be strange to urge that we never do so. We don't conjecture that there is exactly zero correlation between people's impressions of their associates' overall morality and their associates' actual overall morality. Moreover, ordinary moral character

judgments are based on morally relevant behaviors that happen in the context of existing interpersonal relationships. Thus, irrespective of the degree to which people's ordinary moral character judgments track a person's objective overall moral goodness, such judgments can tell us what we need to know to decide whether to date, befriend, hire, or avoid a person (Westra, 2022).

Scientists, too, sometimes reasonably act and theorize based on overall assessments of targets' morality. Noisy measures—as long as they are not entirely inverted or devoid of signal—can be useful in drawing conclusions at the aggregate level or in cases where there's good reason to think the particular sources of inaccuracy are limited or matched between comparisons. However, there are certainly also occasions in which group-level conclusions might be sufficiently harmful or offensive (e.g., dubious claims about civic dishonesty among Chinese people; Cohn et al., 2019; Yang et al., 2023) that they should be avoided unless they pass a very high methodological bar. Moreover, considering the potential influence and seeming objectivity of scientifically supported measures of morality, the stakes can be higher—especially if the results are widely publicized, inform policy decisions, or inform people's self-conceptions. Consequently, higher standards of accuracy can often reasonably be demanded for scientific purposes.

Future Directions

We emphasize that although we think the measurement of morality involves a uniquely challenging array of conceptual and methodological difficulties, we do not recommend abandoning the measurement of specific aspects of morality or even, for limited purposes with appropriate caution, attempts to measure a person's general morality. Some—but not all—of the issues we have raised are empirical questions (e.g., How worried should we be about rating

inversions?). As validity evidence accumulates for both existing and future measures, it might turn out that some measures of general morality will be sufficiently accurate for certain research purposes (especially if the aims are modest, carefully delineated, and confined to group-level inferences). Towards this end, advances in philosophical and psychological research will inform the development of more valid approaches to moral measurement and our understanding of how to minimize the ethical risks.

Moral philosophers might look for areas of consensus and disagreement among different background theories from philosophical and religious traditions across many cultures (e.g., Global Ethic Foundation, 2024; Peterson & Seligman, 2004). This will help researchers understand which aspects of morality are relatively uncontroversial, which aspects need to be interpreted in a relativized way, and the overall degree of moral convergence. For example, if there were some set of virtues (e.g., compassion, honesty, fairness, gratitude, self-control, humility) that comprise a common core of 80% of moral ideas across cultures (with only 20% of moral ideas being cross-culturally variable), this might increase the viability of a cross-culturally applicable moralometer. With enough open communication between people who begin with divergent views, one might optimistically hope for increasing convergence about what is morally good in the long term (Kumar & Campbell, 2022; Parfit, 2011; Pinker, 2018). Even if broad consensus remains elusive, philosophers could further explore, for example, what constitutes virtue in a virtue ethical framework, which rules should govern a deontological framework, and what maximizes good consequences in a consequentialist framework, so that framework-relative measurement can be improved.

We have cautioned against the application of moralometers to draw inferences about particular individuals, especially in high-stakes contexts. However, there could arguably be high-

stakes contexts in which applying a “good enough” (or “better than nothing”) moralometer to an individual is appropriate, perhaps as one factor among many evaluating political candidates or high-risk individuals for monitoring. Ethical work remains to be done exploring the conditions under which it is defensible to reach conclusions about individuals’ or groups’ morality based on imperfect measures with a scientific imprimatur (Mau, 2019). It is also worth exploring the possible ethical downsides of measuring people’s general morality, even with a hypothetical perfect measure. For example, if people who are accurately judged to have poor moral character are excluded from opportunities before they actually perform any relevant bad actions, that raises concerns about the possible injustice of “pre-punishment” (New, 1992; Smilansky, 1994). A case might also potentially be made that one’s moral character is a private matter that shouldn’t be measured without consent.

Future psychological work can help better establish the extent to which a person’s standing on one virtue is diagnostic of their standing on other virtues (Jackson et al., 2023; Landy & Bartels, 2018). This information will help determine whether researchers could rely on one or two particularly central features (e.g., selfishness, compassion) to draw some general conclusions about a person’s moral goodness if measurement resources are constrained. For example, Moshagen and colleagues (2018) have proposed that various “dark” traits (e.g., psychopathy, sadism, vindictiveness) share a common core (D) that reflects “the tendency to maximize one’s own utility at the expense of others” (p. 659). However, it is unclear whether D is adequate for explaining virtuous and supererogatory behaviors (as opposed to only morally questionable behaviors). Understanding the discrepancy between actual and perceived psychological moral unity could also reveal patterns of bias in people’s moral judgments. For example, if people believe that compassion, honesty, and loyalty are strongly correlated when

they are not, people might overweight information about one dimension (e.g., compassion) when judging a person's standing on another dimension (e.g., honesty).

Given the relative feasibility of self-report and reputation-based measures, there is a critical need to develop methods to separate out evaluation from substance in person perception (Leising et al., 2015). For example, researchers have so far explored the effectiveness of euphemistic or dysphemic framings (that make a virtuous behavior sound less virtuous, and vice versa; Meindl et al., 2015; see also Peabody, 1967) or partialling out an evaluativeness factor (Pringle et al., 2024). Another priority is development of unobtrusive behavioral measures that cannot easily be faked (e.g., predictions based on digital footprints; Ebert et al., 2021; Harari et al., 2016; Lewis et al., 2008; Park et al., 2015), bearing in mind that such measures must be consistent with ethical standards of privacy and consent (Matz et al., 2022).

Conclusion

It would be difficult to develop a conceptually sound and highly methodologically valid moralometer that gives an accurate reading of a person's overall moral goodness. However, it doesn't follow that we should stop assessing moral psychological phenomena. Our goal in this paper has been to highlight the unique, value-laden complexities of measuring how morally good (vs. bad) an individual person is, outlining steps that could improve the validity of moral measures and their more accurate interpretation, while encouraging epistemic humility and caution. We are optimistic that by working together, philosophers and psychologists will continue to make conceptual and methodological progress towards improving the measurement of people's moral goodness. Such efforts are unlikely to result in a moralometer that is accurate enough to justify broad, individual-level conclusions about general morality. Nevertheless,

incremental advances in the measurement of moral psychological phenomena will still facilitate scientific discovery.

Constraints on Generality, Positionality, and Citations Statements

Constraints on Generality

The conceptual requirements (see Table 1) draw largely on philosophical arguments, informed by our knowledge of both mainstream Western and some non-Western traditions. We also draw on empirical evidence showing that people have different moral judgments between and within cultures. This evidence is based on studies that include participants from a wide range of contexts, including a study of moral values in 25 countries (Atari et al., 2023). When discussing biases in “commonsense morality,” we primarily draw on studies of U.S. participants. However, the main argument that commonsense morality may differ from philosophical conceptions of morality does not depend on the generalizability of these specific illustrative biases. In general, the philosophical and empirical grounding is largely but not exclusively from Anglophone authors and populations.

The methodological requirements (see Table 2) address general issues associated with self-report, reputation-based, behavioral, and biological measures. Many of these arguments are based on issues inherent to each method and thus unlikely to depend on the population being studied, though the severity of the concerns might vary contextually. For example, group or identity-based differences in peoples’ willingness to truthfully report their own or another person's moral character might vary between large, urban societies compared to closer-knit societies, such as small villages or contexts where multiple generations live in the same households. Also in the latter, informants may possess more information about the target’s

relevant characteristics, thereby increasing the validity of reputation-based measures. These differences are likely to be a matter of degree.

Positionality

The first author is a social-personality psychologist who studies the connections between morality and well-being, and the causes and consequences of moral improvement. As a personality psychologist, she assumes that broad traits exist and can be measured to some extent. In addition, given that her substantial research questions depend on being able to measure morality, she is invested in the conclusion that it is possible to measure some aspect of a person's moral character (even if the claim needs to be narrowed to a specific conceptualization of morality, and even if such measurements are noisy). The first author was born in New Zealand, academically trained in Australia and the U.S., and is currently based in the U.S. These are relatively loose (Gelfand et al., 2011), multicultural societies that tend to be relatively open to new ideas, including different moral ideas. The first author's socialization within such cultures and her expertise with moral psychology may have informed her perspective that it is not possible to conceptualize morality in a single, non-controversial way.

The second author is a philosophy professor born in Boston and living in California since age seven. In his philosophical work, he has frequently defended skepticism about self-knowledge, including pessimism about the accuracy of moral self-report. His work is also informed by an interest in classical Chinese philosophy, seen partly through the lens of recent Anglophone interpretations that connect the ancient Chinese tradition with recent empirical moral psychology. Knowledge of classical Chinese philosophy thus partly shapes his sense of cultural variation in moral evaluation. He has also published several empirical explorations of the extent to which philosophical moral argumentation does or (more often) does not influence

people's moral behavior. This empirical work has generally favored ecologically situated behavioral measurement over self-report and laboratory methods, motivated by the concern that self-report and laboratory measures are likely to generate false positives through demand and social desirability, and the current article continues to reflect that perspective.

This article began in early 2021 as a series of pessimistic reflections by the second author. The first author was at the time doing research reliant on self- and informant reports of moral character, and the second author was impressed by her thoroughness in considering the methodological risks and limitations, so he invited her into the project in a partly adversarial collaboration. Though both authors endorse the entire content of this article, the first author began and remains considerably more optimistic than the second author about the prospects and value of valid measurement of broad moral traits.

Citations

For a random sample of 33% of the 269 references included in the original submission of this paper ($n = 89$), we coded the first author's apparent gender (based on their name and other information available on the internet, e.g., photos on faculty websites), the country of their institution (at the time of the publication), and their field (psychology, philosophy, or other). After dropping duplicate first authors (e.g., if we cited more than one publication from the same first author), 78 first authors remained. First authors (17.95% women, 82.05% men) were primarily psychologists (62.82%) and philosophers (24.36%), with some representation of other fields (12.82%). They were based in nine countries: the U.S. (71.79%), England (7.69%), Germany (6.41%), Canada (3.85%), Australia (3.85%), Sweden (2.56%), India (1.28%), the Netherlands (1.28%), and Switzerland (1.28%).

Acknowledgments

For useful discussion thanks to Willem van der Deijl, Sydney Scott, Nora Williams, and Ben Hardin; and audiences at San Raffaele University, Milan; the Cambridge University Philosophy of Science reading group; the Moral Psychology Research Group; the Washington University in St. Louis Philosophy-Neuroscience-Psychology working group. Thanks to Isabel Thielmann, Nicole Casali, Aurélien Allard, and Daniel Leising for detailed comments. Thanks to Tori Trammell for her assistance with the literature search and citations statement.

Funding Statement

The authors did not receive funding for this work.

Conflicts of Interest Statement

The authors declare no conflicts of interest.

References

- Abburi, H., Parikh, P., Chhaya, N., & Varma, V. (2021). Fine-grained multi-label sexism classification using a semi-supervised multi-level neural approach. *Data Science and Engineering*, 6(4), 359–379. <https://doi.org/10.1007/s41019-021-00168-y>
- Alexander, L., & Moore, M. (2021). Deontological ethics. In *The Stanford Encyclopedia of Philosophy* (Winter 2021).
- Alexandrova, A. (2017). *A philosophy for the science of well-being*. Oxford University Press.
- Allport, G., & Odbert, H. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47(1), i–171. <https://doi.org/doi.org/10.1037/h0093360>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anglim, J., Morse, G., De Vries, R. E., MacCann, C., & Marty, A. (2017). Comparing job applicants to non-applicants using an item-level bifactor model on the hexaco personality inventory. *European Journal of Personality*, 31(6), 669–684. <https://doi.org/10.1002/per.2120>
- Antonoplis, S. (2023). Studying socioeconomic status: Conceptual problems and an alternative path forward. *Perspectives on Psychological Science*, 18(2), 275–292. <https://doi.org/10.1177/17456916221093615>
- Arendt, H. (1963). *Eichmann in Jerusalem: A Report on the Banality of Evil*. Viking Press.
- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63(7), 602–614. <https://doi.org/10.1037/0003-066X.63.7.602>

- Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S. T., & Dehghani, M. (2023). Morality beyond the WEIRD: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspp0000470>
- Atwood, M. (1985). *The handmaid's tale*. McClelland & Stewart.
- Awad, E., Dsouza, S., Shariff, A., Rahwan, I., & Bonnefon, J.-F. (2020). Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proceedings of the National Academy of Sciences*, *117*(5), 2332–2337. <https://doi.org/10.1073/pnas.1911517117>
- Barto, D., Bird, C. W., Hamilton, D. A., & Fink, B. C. (2017). The simple video coder: A free tool for efficiently coding social video data. *Behavior Research Methods*, *49*(4), 1563–1568. <https://doi.org/10.3758/s13428-016-0787-0>
- Batson, C. D., Dyck, J. L., Brandt, J. R., Batson, J. G., Powell, A. L., McMaster, M. R., & Griffitt, C. (1988). Five studies testing two new egoistic alternatives to the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, *55*(1), 52–77. <https://doi.org/10.1037/0022-3514.55.1.52>
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. <https://doi.org/10.1111/j.1745-6916.2007.00051.x>
- Berryman, K., Lazar, S. W., & Hohwy, J. (2023). Do contemplative practices make us more moral? *Trends in Cognitive Sciences*, *27*(10), 916–931. <https://doi.org/10.1016/j.tics.2023.07.005>
- Bicchieri, C. (2017). *Norms in the wild*. Oxford University Press.

Bishop, M. A. (2015). *The good life: Unifying the philosophy and psychology of well-being*. Oxford University Press.

Biswas-Diener, R., & Diener, E. (2006). The subjective well-being of the homeless, and lessons for happiness. *Social Indicators Research*, 76(2), 185–205.
<https://doi.org/10.1007/s11205-005-8671-9>

Bleidorn, W., & Denissen, J. J. A. (2015). Virtues in action – the new look of character traits. *British Journal of Psychology*, 106(4), 700–723. <https://doi.org/10.1111/bjop.12117>

Block, J. (1993). Studying personality the long way. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development* (pp. 9–41). American Psychological Association. <https://doi.org/10.1037/10127-018>

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. <https://doi.org/10.1016/j.jrp.2015.12.003>

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061–1071. <https://doi.org/10.1037/0033-295x.111.4.1061>

Briggs, D. C., Maul, A., & McGrane, J. A. (forthcoming). On the nature of measurement. In L. Cook & M. Pitoniak (Eds.), *Educational Measurement* (5th ed.).

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.

Carlson, R. W., & Crockett, M. (2024, September 13). The pitfalls of pay-to-play morality. <https://doi.org/10.31234/osf.io/ycdfz>

- Carlson, R. W., Maréchal, M. A., Oud, B., Fehr, E., & Crockett, M. J. (2020). Motivated misremembering of selfish decisions. *Nature Communications*, *11*(1), Article 1. <https://doi.org/10.1038/s41467-020-15602-4>
- Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *The Journal of Abnormal and Social Psychology*, *38*(4), 476–506. <https://doi.org/doi.org/10.1037/h0054116>
- Caviola, L., Althaus, D., Mogensen, A. L., & Goodwin, G. P. (2022). Population ethical intuitions. *Cognition*, *218*, 104941. <https://doi.org/10.1016/j.cognition.2021.104941>
- Caviola, L., Schubert, S., & Greene, J. D. (2021). The psychology of (in)effective altruism. *Trends in Cognitive Sciences*, *25*(7), 596–607. <https://doi.org/10.1016/j.tics.2021.03.015>
- Chang, R. (2015). Value incomparability and incommensurability. In I. Hirose & J. Olson (Eds.), *The Oxford Handbook of Value Theory* (pp. 205–224). Oxford University Press.
- Chappell, R. Y., & Meissner, D. (2023). Theories of well-being. In R. Y. Chappell, D. Meissner, & W. MacAskill (Eds.), *An Introduction to Utilitarianism*. <https://www.utilitarianism.net/theories-of-wellbeing>
- Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2020). Lying to appear honest. *Journal of Experimental Psychology: General*, *149*(9), 1719–1735. <https://doi.org/10.1037/xge0000737>
- Christiansen, C. E. (2023). Does fiction reading make us better people? Empathy and morality in a literary empowerment programme. *Ethnos*, *88*(5), 994–1013. <https://doi.org/10.1080/00141844.2021.2007158>
- Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika*, *72*(2), 263–268. <https://doi.org/10.1007/s11336-006-1560-6>

- Cohen, T. R., & Morse, L. (2014). Moral character: What it is and what it does. *Research in Organizational Behavior*, 34, 43–61. <https://doi.org/10.1016/j.riob.2014.08.003>
- Cohen, T. R., Panter, A. T., & Turan, N. (2012). Guilt proneness and moral character. *Current Directions in Psychological Science*, 21(5), 355–359. <https://doi.org/10.1177/0963721412454874>
- Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2014). Moral character in the workplace. *Journal of Personality and Social Psychology*, 107(5), 943–963. <https://doi.org/10.1037/a0037245>
- Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science*, 365(6448), 70–73. <https://doi.org/10.1126/science.aau8712>
- Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment, Vol. 1. Theoretical foundations and research validation; Vol. 2. Standard issue scoring manual*. Cambridge University Press.
- Collier-Spruel, L., Hawkins, A., Jayawickreme, E., Fleeson, W., & Furr, R. M. (2019). Relativism or tolerance? Defining, assessing, connecting, and distinguishing two moral personality features with prominent roles in modern societies. *Journal of Personality*, 87(6), 1170–1188. <https://doi.org/10.1111/jopy.12466>
- 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. <https://doi.org/10.1037/a0021212>
- Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine*, 74(4), 327–337. <https://doi.org/10.1097/PSY.0b013e3182546f18>

- Creemers, R. (2018). *China's social credit system: An evolving practice of control* (SSRN Scholarly Paper 3175792). <https://doi.org/10.2139/ssrn.3175792>
- Crimston, C. R., Bain, P. G., Hornsey, M. J., & Bastian, B. (2016). Moral expansiveness: Examining variability in the extension of the moral world. *Journal of Personality and Social Psychology, 111*(4), 636–653. <https://doi.org/10.1037/pspp0000086>
- Crisp, R. (2021). Well-being. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2021/entries/well-being/>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.
- Dai, Z., Galeotti, F., & Villeval, M. C. (2018). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science, 64*(3), 1081–1100. <https://doi.org/10.1287/mnsc.2016.2616>
- Dancy, J. (2017). Moral particularism. In *The Stanford Encyclopedia of Philosophy* (Winter 2017). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/ENTRIES/moral-particularism/>
- Davidov, M., Zahn-Waxler, C., Roth-Hanania, R., & Knafo, A. (2013). Concern for others in the first year of life: Theory, evidence, and avenues for research. *Child Development Perspectives, 7*(2), 126–131. <https://doi.org/10.1111/cdep.12028>
- DeYoung, C. G. (2015). Cybernetic big five theory. *Journal of Research in Personality, 56*, 33–58. <https://doi.org/10.1016/j.jrp.2014.07.004>
- DeYoung, C. G., Beaty, R. E., Genç, E., Latzman, R. D., Passamonti, L., Servaas, M. N., Shackman, A. J., Smillie, L. D., Spreng, R. N., Viding, E., & Wacker, J. (2022).

- Personality neuroscience: An emerging field with bright prospects. *Personality Science*, 3, e7269. <https://doi.org/10.5964/ps.7269>
- DeYoung, C. G., & Tiberius, V. (2023). Value fulfillment from a cybernetic perspective: A new psychological theory of well-being. *Personality and Social Psychology Review*, 27(1), 3–27. <https://doi.org/10.1177/10888683221083777>
- Diener, E., Oishi, S., & Tay, L. (2018). Advances in subjective well-being research. *Nature Human Behaviour*, 2(4), Article 4. <https://doi.org/10.1038/s41562-018-0307-6>
- Disabato, D. J., Goodman, F. R., Kashdan, T. B., Short, J. L., & Jarden, A. (2016). Different types of well-being? A cross-cultural examination of hedonic and eudaimonic well-being. *Psychological Assessment*, 28(5), 471–482. <https://doi.org/10.1037/pas0000209>
- Dodge, K. A., Bierman, K. L., Coie, J. D., Greenberg, M. T., Lochman, J. E., McMahon, R. J., & Pinderhughes, E. E. (2015). Impact of early intervention on psychopathology, crime, and well-being at age 25. *American Journal of Psychiatry*, 172(1), 59–70. <https://doi.org/10.1176/appi.ajp.2014.13060786>
- Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. Cambridge University Press.
- Doris, J. M. (2022). The future of character. In J. M. Doris (Ed.), *Character Trouble: Undisciplined Essays on Moral Agency and Personality*. Oxford University Press. <https://doi.org/10.1093/oso/9780198719601.003.0011>
- Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. E. (2015). Political diversity will improve social psychological science. *Behavioral and Brain Sciences*, 38, e130. <https://doi.org/10.1017/S0140525X14000430>

- Earp, B. D., McLoughlin, K. L., Monrad, J. T., Clark, M. S., & Crockett, M. J. (2021). How social relationships shape moral wrongness judgments. *Nature Communications*, *12*(1), Article 1. <https://doi.org/10.1038/s41467-021-26067-4>
- Ebert, T., Götz, F. M., Gladstone, J. J., Müller, S. R., & Matz, S. C. (2021). Spending reflects not only who we are but also who we are around: The joint effects of individual and geographic personality on consumption. *Journal of Personality and Social Psychology*, *121*(2), 378–393. <https://doi.org/10.1037/pspp0000344>
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174. <https://doi.org/10.1037/1082-989X.5.2.155>
- Enriquez, J. (2021). *Right/wrong: How technology transforms our ethics*. MIT Press.
- Everett, J. A. C., Faber, N. S., Savulescu, J., & Crockett, M. J. (2018). The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *Journal of Experimental Social Psychology*, *79*, 200–216. <https://doi.org/10.1016/j.jesp.2018.07.004>
- Flanagan, O. (1991). *Varieties of moral personality*. Cambridge, MA: Harvard University Press.
- Fleeson, W., Furr, R. M., Jayawickreme, E., Luke, D., Prentice, M., Reynolds, C. J., & Parham, A. H. (2023). Consensus, controversy, and chaos in the attribution of characteristics to the morally exceptional. *Journal of Personality*. <https://doi.org/10.1111/jopy.12867>
- Fleeson, W., Furr, R. M., Jayawickreme, E., Helzer, E. G., Hartley, A. G., & Meindl, P. (2015). Personality science and the foundations of character. In C. B. Miller, R. M. Furr, A. Knobel, & W. Fleeson (Eds.), *Character: New directions from philosophy, psychology, and theology* (pp. 41–71).

- 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(4), 178–191. <https://doi.org/10.1111/spc3.12094>
- Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, 97(6), 1097–1114. <https://doi.org/10.1037/a0016786>
- Fowers, B. J. (2022). Social science as an inherently moral endeavor. *Journal of Moral Education*, 51(1), 35–46. <https://doi.org/10.1080/03057240.2020.1781069>
- Fowers, B. J., Carroll, J. S., Leonhardt, N. D., & Cokelet, B. (2021). The emerging science of virtue. *Perspectives on Psychological Science*, 16(1), 118–147. <https://doi.org/10.1177/1745691620924473>
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102(4), 652–670. Scopus. <https://doi.org/10.1037/0033-295X.102.4.652>
- Furr, M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, 23(5), 369–401. <https://doi.org/10.1002/per.724>
- Furr, M., Jayawickreme, E., & Santos, C. (2021). *Truthful communication scale (TCS): Conceptual basis & psychometric properties*.
- Furr, M. R., Prentice, M., Hawkins Parham, A., & Jayawickreme, E. (2022). Development and validation of the Moral Character Questionnaire. *Journal of Research in Personality*, 98, 104228. <https://doi.org/10.1016/j.jrp.2022.104228>

- Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliaich, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Subirats Ferrer, M., Fischlmayr, I. C., ... Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*(6033), 1100–1104. <https://doi.org/10.1126/science.1197754>
- Ginges, J., Atran, S., Medin, D., & Shikaki, K. (2007). Sacred bounds on rational resolution of violent political conflict. *Proceedings of the National Academy of Sciences*, *104*(18), 7357–7360. <https://doi.org/10.1073/pnas.0701768104>
- Global Ethic Foundation. (2024, April 22). Stiftung Weltethos. <https://www.weltethos.org/en/>
- Goodman, F. R., Disabato, D. J., Kashdan, T. B., & Kauffman, S. B. (2018). Measuring well-being: A comparison of subjective well-being and PERMA. *The Journal of Positive Psychology*, *13*(4), 321–332. <https://doi.org/10.1080/17439760.2017.1388434>
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–168. <https://doi.org/10.1037/a0034726>
- Gowans, C. W. (2021). *Self-cultivation philosophies in ancient India, Greece and China*. Oxford University Press.
- Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, *8*, 125–130. <https://doi.org/10.1016/j.copsyc.2015.09.007>
- Greene, J. (2014). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin.

- Grubbs, J. B., Perry, S. L., Wilt, J. A., & Reid, R. C. (2019). Pornography problems due to moral incongruence: An integrative model with a systematic review and meta-analysis. *Archives of Sexual Behavior, 48*(2), 397–415. <https://doi.org/10.1007/s10508-018-1248-x>
- Hampson, S. E., Goldberg, L. R., & John, O. P. (1987). Category-breadth and social-desirability values for 573 personality terms. *European Journal of Personality, 1*(4), 241–258. <https://doi.org/10.1002/per.2410010405>
- Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science, 11*(6), 838–854. <https://doi.org/10.1177/1745691616650285>
- Hare, R. D. (1998). The Hare PCL-R: Some issues concerning its use and misuse. *Legal and Criminological Psychology, 3*(1), 99–119. <https://doi.org/10.1111/j.2044-8333.1998.tb00353.x>
- Hartley, A. G., Furr, R. M., Helzer, E. G., Jayawickreme, E., Velasquez, K. R., & Fleeson, W. (2016). Morality's centrality to liking, respecting, and understanding others. *Social Psychological and Personality Science, 7*(7), 648–657. <https://doi.org/10.1177/1948550616655359>
- Hausman, D. M. (2015). *Valuing health: Well-being, freedom, and suffering*. Oxford University Press.
- Haybron, D. M. (2008). Philosophy and the science of subjective well-being. In *The science of subjective well-being* (pp. 17–43). Guilford Press.

- Hayes, A. F., & Dunning, D. (1997). Construal processes and trait ambiguity: Implications for self-peer agreement in personality judgment. *Journal of Personality and Social Psychology*, 72(3), 664–677. <https://doi.org/10.1037/0022-3514.72.3.664>
- Helzer, E. G., Cohen, T. R., Kim, Y., Iorio, A., & Aven, B. (2024). Moral beacons: Understanding moral character and moral influence. *Journal of Personality*. <https://doi.org/10.1111/jopy.12865>
- 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. <https://doi.org/10.1177/0146167214554957>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hester, N., & Gray, K. (2020). The moral psychology of raceless, genderless strangers. *Perspectives on Psychological Science*, 15(2), 216–230. <https://doi.org/10.1177/1745691619885840>
- Hitokoto, H., & Uchida, Y. (2015). Interdependent happiness: Theoretical importance and measurement validity. *Journal of Happiness Studies*, 16(1), 211–239. <https://doi.org/10.1007/s10902-014-9505-8>
- Hofstee, W. K. B. (1994). *Who should own the definition of personality?* 8(3), 149–162. <https://doi.org/10.1002/per.2410080302>
- Hopwood, C. J., Bleidorn, W., Schwaba, T., & Chen, S. (2020). Health, environmental, and animal rights motives for vegetarian eating. *PLOS ONE*, 15(4), e0230609. <https://doi.org/10.1371/journal.pone.0230609>

- Hursthouse, R., & Pettigrove, G. (2023). Virtue ethics. In *The Stanford Encyclopedia of Philosophy* (Fall 2023). Metaphysics Research Lab, Stanford University.
<https://plato.stanford.edu/archives/fall2023/entries/ethics-virtue>
- Huxley, A. (1932). *Brave new world*. Chatto & Windus.
- Inbar, Y., & Lammers, J. (2012). Political diversity in social and personality psychology. *Perspectives on Psychological Science*, 7(5), 496–503.
<https://doi.org/10.1177/1745691612448792>
- Jackson, J. C., Halberstadt, J., Takezawa, M., Liew, K., Smith, K., Apicella, C., & Gray, K. (2023). Generalized morality culturally evolves as an adaptive heuristic in large social networks. *Journal of Personality and Social Psychology*, 125(6), 1207–1238.
<https://doi.org/10.1037/pspa0000358>
- Jaeger, B., & van Vugt, M. (2022). Psychological barriers to effective altruism: An evolutionary perspective. *Current Opinion in Psychology*, 44, 130–134.
<https://doi.org/10.1016/j.copsyc.2021.09.008>
- Jenni, K., & Loewenstein, G. (1997). Explaining the identifiable victim effect. *Journal of Risk and Uncertainty*, 14(3), 235–257. <https://doi.org/10.1023/A:1007740225484>
- Jiang, L., Hwang, J. D., Bhagavatula, C., Bras, R. L., Liang, J., Dodge, J., Sakaguchi, K., Forbes, M., Borchardt, J., Gabriel, S., Tsvetkov, Y., Etzioni, O., Sap, M., Rini, R., & Choi, Y. (2022). *Can Machines Learn Morality? The Delphi Experiment* (arXiv:2110.07574). arXiv. <https://doi.org/10.48550/arXiv.2110.07574>
- Kagan, S. (1989). *The limits of morality*. Oxford University Press.

- Kahane, G., Everett, J. A. C., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, *125*(2), 131–164. <https://doi.org/10.1037/rev0000093>
- Kaufman, S. B. (2014). *Will the real introverts please stand up?* Scientific American Blog Network. <https://blogs.scientificamerican.com/beautiful-minds/will-the-real-introverts-please-stand-up/>
- Kennedy, B., Atari, M., Davani, A. M., Yeh, L., Omrani, A., Kim, Y., Coombs, K., Havaladar, S., Portillo-Wightman, G., Gonzalez, E., Hoover, J., Azatian, A., Hussain, A., Lara, A., Cardenas, G., Omary, A., Park, C., Wang, X., Wijaya, C., ... Dehghani, M. (2022). Introducing the gab hate corpus: Defining and applying hate-based rhetoric to social media posts at scale. *Language Resources and Evaluation*, *56*(1), 79–108. <https://doi.org/10.1007/s10579-021-09569-x>
- Kennedy, B., Atari, M., Mostafazadeh Davani, A., Hoover, J., Omrani, A., Graham, J., & Dehghani, M. (2021). Moral concerns are differentially observable in language. *Cognition*, *212*, 104696. <https://doi.org/10.1016/j.cognition.2021.104696>
- Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, *98*(2), 155–163. <https://doi.org/10.1037/0033-295X.98.2.155>
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, *8*(3), 265–280. https://doi.org/10.1207/s15327957pspr0803_3
- Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, *43*(1), 23–34. <https://doi.org/10.1037/0003-066X.43.1.23>

- Kim, H., Di Domenico, S. I., & Connelly, B. S. (2019). Self–other agreement in personality reports: A meta-analytic comparison of self- and informant-report means. *Psychological Science, 30*(1), 129–138. <https://doi.org/10.1177/0956797618810000>
- Klein, N., & Epley, N. (2017). Less evil than you: Bounded self-righteousness in character inferences, emotional reactions, and behavioral extremes. *Personality and Social Psychology Bulletin, 43*(8), 1202–1212. <https://doi.org/10.1177/0146167217711918>
- Kohlberg, L., Levine, C., & Hewer, A. (1983). Moral stages: A current formulation and a response to critics. *Contributions to Human Development, 10*, 174–174.
- Kröll, M., & Rustagi, D. (2016). Reputation, honesty, and cheating in informal milk markets in india. In *134_857756613* [Working Paper]. <https://doi.org/10.2139/ssrn.2982365>
- Krueger, R. F., & Johnson, W. (2021). Behavioral genetics and personality: Ongoing efforts to integrate nature and nurture. In *Handbook of personality: Theory and research, 4th ed* (pp. 217–241). The Guilford Press.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one’s own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Kumar, V., & Campbell, R. (2022). *A better ape*. Oxford University Press.
- Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomezsko, D., Greenwald, A. G., & Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist, 74*(5), 569–586. <https://doi.org/10.1037/amp0000364>

- Landy, J. F., & Bartels, D. M. (2018). An empirically-derived taxonomy of moral concepts. *Journal of Experimental Psychology: General*, *147*(11), 1748–1761.
<https://doi.org/10.1037/xge0000404>
- Law, K. F., Campbell, D., & Gaesser, B. (2022). Biased benevolence: The perceived morality of effective altruism across social distance. *Personality and Social Psychology Bulletin*, *48*(3), 426–444. <https://doi.org/10.1177/01461672211002773>
- Law, K. F., Syropoulos, S., Coleman, M., Gainsburg, I., & O'Connor, B. B. (2023). *Moral future-thinking: Does the moral circle stand the test of time?* PsyArXiv.
<https://doi.org/10.31234/osf.io/c75ny>
- Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment*, *25*(5), 543–556. <https://doi.org/10.1177/1073191116659134>
- Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, *98*(4), 668–682.
<https://doi.org/10.1037/a0018771>
- Leising, D., Scherbaum, S., Locke, K. D., & Zimmermann, J. (2015). A model of “substance” and “evaluation” in person judgments. *Journal of Research in Personality*, *57*, 61–71.
<https://doi.org/10.1016/j.jrp.2015.04.002>
- Lenman, J. (2000). Consequentialism and cluelessness. *Philosophy & Public Affairs*, *29*(4), 342–370. <https://doi.org/10.1111/j.1088-4963.2000.00342.x>
- Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, *30*(4), 330–342. <https://doi.org/10.1016/j.socnet.2008.07.002>

- Liang, F., Das, V., Kostyuk, N., & Hussain, M. M. (2018). Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. *Policy & Internet, 10*(4), 415–453. <https://doi.org/10.1002/poi3.183>
- Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2018). The role of a “common is moral” heuristic in the stability and change of moral norms. *Journal of Experimental Psychology: General, 147*(2), 228–242. <https://doi.org/10.1037/xge0000365>
- Lucas, R. E., & Donnellan, M. B. (2009). If the person–situation debate is really over, why does it still generate so much negative affect? *Journal of Research in Personality, 43*(2), 146–149. <https://doi.org/10.1016/j.jrp.2009.02.009>
- Lukaszewski, A. W., & Roney, J. R. (2010). Kind toward whom? Mate preferences for personality traits are target specific. *Evolution and Human Behavior, 31*(1), 29–38. <https://doi.org/10.1016/j.evolhumbehav.2009.06.008>
- Margolis, S., Schwitzgebel, E., Ozer, D. J., & Lyubomirsky, S. (2021). Empirical relationships among five types of well-being. In *Measuring Well-being: Interdisciplinary Perspectives from the Social Sciences and the Humanities*. Oxford University Press.
- Mast, M. S., Gatica-Perez, D., Frauendorfer, D., Nguyen, L., & Choudhury, T. (2015). Social sensing for psychology: Automated interpersonal behavior assessment. *Current Directions in Psychological Science, 24*(2), 154–160. <https://doi.org/10.1177/0963721414560811>
- Mastroianni, A. M., & Gilbert, D. T. (2023). The illusion of moral decline. *Nature, 618*(7966), Article 7966. <https://doi.org/10.1038/s41586-023-06137-x>

- Matz, S. C., Appel, R. E., & Croll, B. (2022). Privacy and ethics in the age of Big Data. In *The psychology of technology: Social science research in the age of Big Data* (pp. 379–420). American Psychological Association. <https://doi.org/10.1037/0000290-012>
- Mau, S. (2019). *The metric society: On the quantification of the social*. John Wiley & Sons.
- Mehl, M. R. (2017). The electronically activated recorder (EAR): A method for the naturalistic observation of daily social behavior. *Current Directions in Psychological Science*, 26(2), 184–190. <https://doi.org/10.1177/0963721416680611>
- Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The electronically activated recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*, 33(4), 517–523. <https://doi.org/10.3758/BF03195410>
- Meindl, P., & Graham, J. (2014). Know thy participant: The trouble with nomothetic assumptions in moral psychology. In *Advances in Experimental Moral Psychology* (pp. 233–252). A&C Black.
- Meindl, P., Jayawickreme, E., Furr, R. M., & Fleeson, W. (2015). A foundation beam for studying morality from a personological point of view: Are individual differences in moral behaviors and thoughts consistent? *Journal of Research in Personality*, 59, 81–92. <https://doi.org/10.1016/j.jrp.2015.09.005>
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45, 35–44.
- Miller, C. B. (2017). Honesty. In W. Sinnott-Armstrong & C.B. Miller (Eds.), *Moral psychology: Virtue and character* (pp. 237–273). Boston Review. <https://psycnet.apa.org/record/2017-17609-018>

- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Mischel, W., Coates, B., & Raskoff, A. (1968). Effects of success and failure on self-gratification. *Journal of Personality and Social Psychology*, *10*(4), 381–390.
<https://doi.org/10.1037/h0026800>
- Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656–688. <https://doi.org/10.1037/rev0000111>
- Möttus, R. (2022). What correlations mean for individual people: A tutorial for researchers, students and the public. *Personality Science*, *3*, 1–27. <https://doi.org/10.5964/ps.7467>
- New, C. (1992). Time and punishment. *Analysis*, *52*(1), 35–40. <https://doi.org/10.2307/3328880>
- Norden, B. W. V., & Ivanhoe, P. J. (2023). *Readings in classical chinese philosophy*. Hackett Publishing.
- Norman, W. T. (1967). *2800 Personality trait descriptors—Normative operating characteristics for a university population*. <https://eric.ed.gov/?id=ed014738>
- Nussbaum, M. (1997). *Cultivating humanity*. Harvard University Press.
- Oishi, S., Diener, E., Suh, E., & Lucas, R. E. (1999). Value as a moderator in subjective well-being. *Journal of Personality*, *67*(1), 157–184. <https://doi.org/10.1111/1467-6494.00051>
- Oishi, S., Graham, J., Kesebir, S., & Galinha, I. C. (2013). Concepts of happiness across time and cultures. *Personality and Social Psychology Bulletin*, *39*(5), 559–577.
<https://doi.org/10.1177/0146167213480042>
- Olaru, G., Schroeders, U., Wilhelm, O., & Ostendorf, F. (2019). ‘Grandpa, do you like roller coasters?’: Identifying age-appropriate personality indicators. *European Journal of Personality*, *33*(3), 264–278. <https://doi.org/10.1002/per.2185>
- Orwell, G. (1949). *Nineteen eighty-four*. Secker & Warburg.

- Paluck, E. L., & Shepherd, H. (2012). The salience of social referents: A field experiment on collective norms and harassment behavior in a school social network. *Journal of Personality And Social Psychology, 103*(6), 899–915. <https://doi.org/10.1037/a0030015>
- Parfit, D. (1984). *Reasons and persons*. OUP Oxford.
- Parfit, D. (2011). *On what matters*. OUP Oxford.
- Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology, 108*(6), 934–952. <https://doi.org/10.1037/pspp0000020>
- Pascual-Ezama, D., Prelec, D., Muñoz, A., & Gil-Gómez de Liaño, B. (2020). Cheaters, liars, or both? A New classification of dishonesty profiles. *Psychological Science, 31*(9), 1097–1106. <https://doi.org/10.1177/0956797620929634>
- Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*(6), 1025–1060. <https://doi.org/10.1111/1467-6494.00041>
- Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology, 7*(4, Pt.2), 1–18. <https://doi.org/10.1037/h0025230>
- Persson, I., & Savulescu, J. (2012). *Unfit for the future: The need for moral enhancement*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199653645.001.0001>
- Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. American Psychological Association.
- Pinker, S. (2018). *Enlightenment now: The case for reason, science, humanism, and progress*. Penguin UK.

- Pringle, V., Sun, J., & Carlson, E. N. (2024). What is the moral person like? An examination of the shared and unique perspectives on moral character. *Journal of Personality, 92*(3), 697–714. <https://doi.org/10.1111/jopy.12902>
- Prinzing, M. (2021). How to study well-being: A proposal for the integration of philosophy with science. *Review of General Psychology, 25*(2), 152–162. <https://doi.org/10.1177/10892680211002443>
- Prinzing, M. M. (2021). Positive psychology is value-laden—It’s time to embrace it. *The Journal of Positive Psychology, 16*(3), 289–297. <https://doi.org/10.1080/17439760.2020.1716049>
- Purcell, Z. A., & Bonnefon, J.-F. (2023). Humans feel too special for machines to score their morals. *PNAS Nexus, 2*(6), pgad179. <https://doi.org/10.1093/pnasnexus/pgad179>
- Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review, 118*(1), 57–75. <https://doi.org/10.1037/a0021867>
- Ramírez-Esparza, N., Mehl, M. R., Álvarez-Bermúdez, J., & Pennebaker, J. W. (2009). Are Mexicans more or less sociable than Americans? Insights from a naturalistic observation study. *Journal of Research in Personality, 43*(1), 1–7. <https://doi.org/10.1016/j.jrp.2008.09.002>
- Redding, R. E. (2001). Sociopolitical diversity in psychology: The case for pluralism. *American Psychologist, 56*(3), 205–215. <https://doi.org/10.1037/0003-066X.56.3.205>
- Rest, J., Thoma, S., & Edwards, L. (1997). Designing and validating a measure of moral judgment: Stage preference and stage consistency approaches. *Journal of Educational Psychology, 89*(1), 5–28. <https://doi.org/10.1037/0022-0663.89.1.5>

- Roberts, R. C., & West, R. (2020). The virtue of honesty: A conceptual exploration. In C. B. Miller & R. West (Eds.), *Integrity, Honesty, and Truth Seeking*. Oxford University Press.
<https://doi.org/10.1093/oso/9780190666026.003.0004>
- Robinson, B. (2020). I am so humble! In M. Alfano, M. P. Lynch, & A. Tanesini (Eds.), *The Routledge Handbook of the Philosophy of Humility*. London: Routledge.
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. McGraw-Hill.
- Rushton, J. P., Chrisjohn, R. D., & Cynthia Fekken, G. (1981). The altruistic personality and the self-report altruism scale. *Personality and Individual Differences*, 2(4), 293–302.
[https://doi.org/10.1016/0191-8869\(81\)90084-2](https://doi.org/10.1016/0191-8869(81)90084-2)
- Salekin, R. T., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist—Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice*, 3(3), 203–215.
<https://doi.org/10.1111/j.1468-2850.1996.tb00071.x>
- Scanlon, T. M. (1998). *What we owe to each other*. Harvard University Press.
- Schein, C. (2020). The importance of context in moral judgments. *Perspectives on Psychological Science*, 15(2), 207–215. <https://doi.org/10.1177/1745691620904083>
- Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, 22(1), 32–70.
<https://doi.org/10.1177/1088868317698288>
- Schild, C., Lilleholt, L., & Zettler, I. (2021). Behavior in cheating paradigms is linked to overall approval rates of crowdworkers. *Journal of Behavioral Decision Making*, 34(2), 157–166. <https://doi.org/10.1002/bdm.2195>

- Schur, M. (Director). (2016, 2020). *The good place*. 3 Arts Entertainment, Fremulon, Universal Television.
- Schwitzgebel, E. (2019). Aiming for moral mediocrity. *Res Philosophica*, *96*(3), 347–368.
<https://doi.org/10.11612/resphil.1806>
- Schwitzgebel, E. (2024). Repetition and value in an infinite universe. In S. Hetherington (Ed.), *Extreme Philosophy*. New York: Routledge.
- Schwitzgebel, E., Cokelet, B., & Singer, P. (2023). Students eat less meat after studying meat ethics. *Review of Philosophy and Psychology*, *14*(1), 113–138.
<https://doi.org/10.1007/s13164-021-00583-0>
- Schwitzgebel, E., & Rust, J. (2009). The moral behaviour of ethicists: Peer opinion. *Mind*, *118*(472), 1043–1059.
- Schwitzgebel, E., & Rust, J. (2014). The moral behavior of ethics professors: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior. *Philosophical Psychology*, *27*(3), 293–327.
<https://doi.org/10.1080/09515089.2012.727135>
- Sedikides, C., Meek, R., Alicke, M. D., & Taylor, S. (2014). Behind bars but above the bar: Prisoners consider themselves more prosocial than non-prisoners. *British Journal of Social Psychology*, *53*(2), 396–403. <https://doi.org/10.1111/bjso.12060>
- Shafer-Landau, R. (1994). Ethical disagreement, ethical objectivism and moral indeterminacy. *Philosophy and Phenomenological Research*, *54*(2), 331–344.
<https://doi.org/10.2307/2108492>
- Sinnott-Armstrong, W. (2022). Consequentialism. In *The Stanford Encyclopedia of Philosophy* (Winter 2022). Metaphysics Research Lab, Stanford University.

- Sinnott-Armstrong, W., & Wheatley, T. (2012). The disunity of morality and why it matters to philosophy. *The Monist*, *95*(3), 355–377. <https://doi.org/10.5840/monist201295319>
- Sinott-Armstrong, W. (2019). Moral skepticism. In *The Stanford Encyclopedia of Philosophy* (Summer 2019). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2019/entries/skepticism-moral/>
- Smilansky, S. (1994). The time to punish. *Analysis*, *54*(1), 50–53. <https://doi.org/10.2307/3328104>
- Smith, I. H., & Kouchaki, M. (2018). Moral humility: In life and at work. *Research in Organizational Behavior*, *38*, 77–94. <https://doi.org/10.1016/j.riob.2018.12.001>
- Smith, K. M., & Apicella, C. L. (2020). Hadza hunter-gatherers disagree on perceptions of moral character. *Social Psychological and Personality Science*, *11*(5), 616–625. <https://doi.org/10.1177/1948550619865051>
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*(1), 117–143. <https://doi.org/10.1037/pspp0000096>
- Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, *27*(1), 76–105. [https://doi.org/10.1016/0022-1031\(91\)90011-T](https://doi.org/10.1016/0022-1031(91)90011-T)
- Sun, J. (2020). *Eavesdropping on interpersonal behavior in everyday life* [Unpublished doctoral dissertation]. University of California, Davis.
- Sun, J., & Goodwin, G. P. (2020). Do people want to be more moral? *Psychological Science*, *31*(3), 243–257. <https://doi.org/10.1177/0956797619893078>

- Sun, J., Harris, K., & Vazire, S. (2020). Is well-being associated with the quantity and quality of social interactions? *Journal of Personality and Social Psychology*, *119*(6), 1478–1496. <https://doi.org/10.1037/pspp0000272>
- Sun, J., Neufeld, B., Snelgrove, P., & Vazire, S. (2022). Personality evaluated: What do people most like and dislike about themselves and their friends? *Journal of Personality and Social Psychology*, *122*(4), 731–748. <https://doi.org/10.1037/pspp0000388>
- Sun, J., & Smillie, L. D. (2024). Why moral psychology needs personality psychology. *Journal of Personality*, *92*(3), 653–665. <https://doi.org/10.1111/jopy.12919>
- Sun, J., Wilt, J., Meindl, P., Watkins, H. M., & Goodwin, G. P. (2023). How and why people want to be more moral. *Journal of Personality*. <https://doi.org/10.1111/jopy.12812>
- Sun, J., Wu, W., & Goodwin, G. P. (2025). Are moral people happier? Answers from reputation-based measures of moral character. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/pspp0000539>
- Syropoulos, S., Law, K. F., & Young, L. (2023). *Caring for present and future generations alike: Longtermism and moral regard across temporal and social distance*. PsyArXiv. <https://doi.org/10.31234/osf.io/hzwrt>
- Tackman, A. M., Baranski, E. N., Danvers, A. F., Sbarra, D. A., Raison, C. L., Moseley, S. A., Polsinelli, A. J., & Mehl, M. R. (2020). ‘Personality in its natural habitat’ revisited: A pooled, multi-sample examination of the relationships between the Big Five personality traits and daily behaviour and language use. *European Journal of Personality*, *34*(5), 753–776. <https://doi.org/10.1002/per.2283>

- Tang, F., Jang, H., Rauktis, M. B., Musa, D., & Beach, S. (2019). The race paradox in subjective wellbeing among older Americans. *Ageing & Society, 39*(3), 568–589.
<https://doi.org/10.1017/S0144686X17001064>
- 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. <https://doi.org/10.1002/per.2112>
- Tiberius, V. (2004). Cultural differences and philosophical accounts of well-being. *Journal of Happiness Studies, 5*(3), 293–314. <https://doi.org/10.1007/s10902-004-8791-y>
- Tiberius, V., & Haybron, D. M. (2022). Prudential psychology: Theory, method, and measurement. In M. Vargas, & J. M. Doris (Eds.), *The Oxford Handbook of Moral Psychology*. <https://doi.org/10.1093/oxfordhb/9780198871712.013.31>
- Vanderford, M. L. (1989). Vilification and social movements: A case study of pro-life and pro-choice rhetoric. *Quarterly Journal of Speech, 75*(2), 166–182.
<https://doi.org/10.1080/00335638909383870>
- Vazire, 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.
<https://doi.org/10.1037/a0017908>
- von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology, 34*(4), 1607–1622. <https://doi.org/10.1007/s13347-021-00477-0>
- Wacker, J., & Smillie, L. D. (2015). Trait extraversion and dopamine function. *Social and Personality Psychology Compass, 9*(6), 225–238. <https://doi.org/10.1111/spc3.12175>

- Walker, L. J., & Frimer, J. A. (2007). Moral personality of brave and caring exemplars. *Journal of Personality and Social Psychology*, *93*(5), 845–860. <https://doi.org/10.1037/0022-3514.93.5.845>
- Walker, L. J., Frimer, J. A., & Dunlop, W. L. (2010). Varieties of moral personality: Beyond the banality of heroism. *Journal of Personality*, *78*(3), 907–942. <https://doi.org/10.1111/j.1467-6494.2010.00637.x>
- Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311*(5765), 1301–1303. <https://doi.org/10.1126/science.1121448>
- Welch, K. (2007). Black criminal stereotypes and racial profiling. *Journal of Contemporary Criminal Justice*, *23*(3), 276–288. <https://doi.org/10.1177/1043986207306870>
- Wessels, N. M., Zimmermann, J., Biesanz, J. C., & Leising, D. (2020). Differential associations of knowing and liking with accuracy and positivity bias in person perception. *Journal of Personality and Social Psychology*, *118*(1), 149–171. <https://doi.org/10.1037/pspp0000218>
- Westra, E. (2022). In defense of ordinary moral character judgment. *Erkenntnis*, *87*(4), 1461–1479. <https://doi.org/10.1007/s10670-020-00257-w>
- Williams, E. G. (2015). The possibility of an ongoing moral catastrophe. *Ethical Theory and Moral Practice*, *18*(5), 971–982. <https://doi.org/10.1007/s10677-015-9567-7>
- Wilson, A. T. (2018). Honesty as a virtue. *Metaphilosophy*, *49*(3), 262–280. <https://doi.org/10.1111/meta.12303>
- Wong, D. B. (2006). *Natural moralities: A defense of pluralistic relativism*. Oxford University Press.

- Wright, J. C., Warren, M. T., & Snow, N. E. (2020). *Understanding virtue: Theory and measurement*. Oxford University Press.
- Yaden, D. B., & Anderson, D. E. (2021). The psychology of philosophy: Associating philosophical views with psychological traits in professional philosophers. *Philosophical Psychology*, 34(5), 721–755. <https://doi.org/10.1080/09515089.2021.1915972>
- Yang, Q., Zhang, W., Liu, S., Gong, W., Han, Y., Lu, J., Jiang, D., Nie, J., Lyu, X., Liu, R., Jiao, M., Qu, C., Zhang, M., Sun, Y., Zhou, X., & Zhang, Q. (2023). Unraveling controversies over civic honesty measurement: An extended field replication in China. *Proceedings of the National Academy of Sciences*, 120(29), e2213824120. <https://doi.org/10.1073/pnas.2213824120>
- Youyou, W., Stillwell, D., Schwartz, H. A., & Kosinski, M. (2017). Birds of a feather do flock together: Behavior-based personality-assessment method reveals personality similarity among couples and friends. *Psychological Science*, 28(3), 263–404. <https://doi.org/10.1177/0956797616678187>
- Yudkin, D., Goodwin, G., Reece, A., Gray, K., & Bhatia, S. (2023). *A large-scale investigation of everyday moral dilemmas*. PsyArXiv. <https://doi.org/10.31234/osf.io/5pcew>
- Zhao, K., Ferguson, E., & Smillie, L. D. (2017). Individual differences in good manners rather than compassion predict fair allocations of wealth in the dictator game. *Journal of Personality*, 85(2), 244–256. <https://doi.org/10.1111/jopy.12237>
- Zimmerman, M. J. (2015). Moral luck reexamined. In D. Shoemaker (Ed.), *Oxford Studies in Agency and Responsibility: Volume 3* (p. 0). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198744832.003.0008>

Ziv, T., & Sommerville, J. A. (2017). Developmental differences in infants' fairness expectations from 6 to 15 months of age. *Child Development, 88*(6), 1930–1951.

<https://doi.org/10.1111/cdev.12674>

1. Perspectives on the (Dis)unity of Morality

Table S1

Summary of Proposals for and Arguments Against the Unity of Morality

Morality is unified on the basis of...	Argument against
Classical utilitarianism: The resulting balance of pleasure vs. pain.	Other consequentialists appeal to a broader range of goods, and accurately evaluating all relevant consequences might be infeasible
Deontology: A single all-encompassing rule (e.g., Kant's maxim of universalizability).	Other deontologists see morality as involving adherence to multiple possibly conflicting rules or obligations (Stocker, 1992; Tessman, 2014; Williams, 1965).
Aristotelian virtue ethics: You can have one virtue if and only if you have them all (Russell, 2009).	Other virtue ethicists reject the unity-of-virtues approach, e.g., courage doesn't require compassion (Badhwar, 1996).
Psychological components	
Content (e.g., harm to others; Schein & Gray, 2018; a "dark core" of utility maximization along with disutility infliction; Moshagen et al., 2018; "obligatory concerns with others' welfares, rights, fairness, and justice"; Dahl, 2023).	Not all moral judgments are based on harm (Graham et al., 2011); not all moral behaviors involve utility maximization or disutility infliction; not all moral concerns are obligatory (Janoff-Bulman et al., 2009; Urmson, 1958).
Phenomenology: How it feels or seems to make a moral judgment (Horgan & Timmons, 2005).	Different moral judgments are associated with different moral emotions (Rozin et al., 1999)
Force: Moral wrongness is authority-independent, serious, and harm-based (Turiel, 1983).	There are some moral judgments that are not authority-independent, serious, or harm-based (McGraw & Warren, 2010; Sinnott-Armstrong & Wheatley, 2012).
Form: A moral judgment is prescriptive, universalizable, and overriding (Hare, 1981).	Moral judgments take a variety of forms and need not always be prescriptive, universalizable, or overriding (Macintyre, 1957; Malle, 2021; Portmore, 2008).
Function: Morality evolved to solve a particular problem (e.g., limits on sympathy; Warnock, 1971; cooperation; Curry et al., 2019)	It seems likely that different moral judgments serve different functions (Graham et al., 2013).
Brain mechanisms (Moll et al., 2005)	Different moral judgments rely on different neural systems (Sinnott-Armstrong, 2016)

2. The Prospects of a Biological Moralometer

Genetic, neural, and physiological measures—at least as they currently exist—are either too indirect, too narrowly targeted, or too noisy to serve as reliable indicators of individual-level general moral tendencies. First, consider genetic measures. Psychological traits are determined

not only by genes but also by a lifetime of learning, experiences, and environmental influences (Bouchard, 2004). A person who entirely turns around their life, transforming themselves from a morally terrible person into an admirable one, will measure the same on a genetically based moralometer before and after the transformation. Any measure that is incapable of capturing such moral change will necessarily be seriously incomplete. Empirically, the predictive power of polygenic scores has been estimated to range from 2% of the variance in self-reported callous-unemotional traits (Wertz et al., 2018) to 11% of the variance in self-reported empathy (Warrier et al., 2018). As indicators of the theoretical upper limit of a genetic moralometer, the most predictive polygenic score in the behavioral sciences (for a review, see Plomin & von Stumm, 2022) predicts 15% of the variance in tested school performance at age 16 (Allegrini et al., 2019). Even for the straightforward physical phenotype of height—which is both more easily measurable and more highly heritable than most psychological traits (Bouchard, 2004; Silventoinen, 2003)—polygenic scores explain approximately 22% of the variance in observed adult height in the United Kingdom (You et al., 2021). Thus, it seems unlikely that a genetic moralometer could ever achieve the level of accuracy required to draw firm conclusions about an individual person’s morality.

If genetic measures are too distal, could the structure and function of a person’s brain be a valid marker of their morality? The most definitive ($N = 1,105$, preregistered) study to date strongly suggest that relations between personality traits and brain structure (e.g., cortical thickness or surface area across the whole brain or in specific regions of interest) tend to be null-to-small (largest $R^2 = .057$ and mean $R^2 = .003\%$; Hyatt et al., 2022), no matter how they are operationalized. It seems extremely unlikely that we could infer how moral a person is simply based on the size and shape of their brain.

Perhaps differences in brain structure are not as clearly linked to personality as differences in brain function (i.e., when and where brain activity occurs at rest and during specific tasks). Turning to brain function, there is some suggestive evidence that, compared to healthy controls, the right amygdala of altruistic kidney donors is more responsive to fearful facial expressions (Marsh et al., 2014), whereas the reverse is true for psychopaths (for a review, see Blair, 2013). Compared to matched control participants, altruistic kidney donors also show greater self–other overlap in neural representations of pain and threat (Brethel-Haurwitz et al., 2018). Some researchers have also found it possible to predict self-reported levels of two morally-relevant traits—agreeableness (around $r = .20$) and conscientiousness (around $r = .30$)—from resting EEG activity (Jach et al., 2020). However, such effect sizes are too small to draw meaningful conclusions about individual-level morality. For example, an effect size of $d = 0.93$ (Marsh et al., 2014) corresponds to a 64.6% overlap between extraordinary altruists and healthy controls in their right amygdala volume (Magnusson, 2023). Moreover, in a meta-analysis of the associations between trait empathy and neural responses, none of the 15 studies had a sample of more than 30 participants (Lamm et al., 2011). Since large studies typically reveal much smaller correlations than small studies (Carre et al., 2013; Bjørnebekka et al., 2013; Liu et al., 2013), effect size estimates based on such small sample sizes are likely inflated (Allen & DeYoung, 2015). Although investigations could eventually provide insight into the neural mechanisms that underpin morality, for the foreseeable future it is unlikely that such measures could provide accurate assessments of a particular person’s morality.

Other physiological measures seem similarly unpromising. For example, “lie detector” tests based on indicators of autonomic arousal (e.g., heart rate, respiration, and skin conductance) are the closest thing we have to a physiological measure of a person’s honesty. However, the

validity of such tests has long been controversial (Saxe & Ben-Shakhar, 1999) because the measured pattern of physiological reactions is not unique to deception. An honest person could be nervous even when telling the truth, and a dishonest psychopath could “beat” the lie detector because of their decreased fear responses (Lykken, 1978).

3. Conceptual and Methodological Requirements for Measuring Specific Moral Traits

In the body of the article, we argue that it would be difficult to develop a conceptually sound and highly methodologically valid general moralometer. As we note in the article, researchers rarely aim to measure general morality, perhaps in light of the concerns we present systematically there. In contrast, there has been relatively more interest in measuring specific moral traits. The question arises to what extent our concerns apply to the measurement of specific moral traits. We suggest that similar conceptual and methodological concerns do apply to a large extent, though less severely. To illustrate this, we consider how these concerns apply to two prototypical moral traits—compassion and honesty.

Conceptual Requirements for Measuring Specific Moral Traits

To measure a specific moral dimension such as compassion or honesty: (1) There must be general facts about people’s compassion or honesty; (2) The measure must correctly identify which characteristics are (un)compassionate or (dis)honest; (3) The measure must correctly weigh different components of compassion or honesty against each other; and (4) the measure must apply clearly and consistently across people and time. As with general morality, we could rely on judges’ subjective, flexible understanding of what it means to be compassionate or honest and how to weigh their various facets, or use a measure that applies the same prespecified, fixed

criteria. Table S2 expresses our sense of the relative seriousness of the conceptual challenges for general morality vs. these two example traits.

Table S2

Conceptual Requirements for Constructing Flexible vs. Fixed Measures of General Morality (M) vs. Compassion (C) or Honesty (H)

Conceptual requirement	Flexible		Fixed	
	M	C/H	M	C/H
1. There must be general moral facts about a person's overall morality/compassion/honesty.				
Realism: There must be facts about what is (im)moral/(un)compassionate/(dis)honest; <i>and</i>	✓	✓	✓	✓
Universalism: The same general things must be (im)moral/(un)compassionate/(dis)honest for different people or in different groups; <i>and</i>	!	!	!!	!
Generalism: What is (im)moral/(un)compassionate/(dis)honest must not depend on highly particular features of specific situations.	!	!	!!	!
2. The measure must correctly identify which characteristics are (im)moral/(un)compassionate/(dis)honest.				
Judges' idiosyncratic moral judgments are correct; <i>or</i>	!!	!	-	-
Commonsense ethics, as operationalized by the researchers, is correct; <i>or</i>	-	-	!!	!
The favored expert ethical framework is correct.	-	-	!!	!
3. The measure must correctly weigh different components of morality/compassion/honesty against each other.				
Unity: General morality/compassion/honesty must be a coherent construct; <i>and</i>	!!	!	!!	!
Commensurability: There must be a common "currency" in which components of morality/compassion/honesty can be appropriately compared.	!!	!	!!	!
4. The measure must apply clearly and consistently across people and time.				
Transparency: It must be clear what is being measured; <i>and</i>	!!	!	✓	✓
Equivalence:				
A fixed measure must capture the same thoughts, feelings, and/or behaviors across people; <i>and</i>	-	-	✓	✓
A fixed measure must have the same moral significance across people; <i>or</i>	-	-	!!	!
For flexible measures, judges must use the same criteria to judge a person's morality/compassion/honesty; <i>and, depending on the aim</i>	!!	!	-	-
The measure must be psychologically relevant.	!	✓	!!	!

Note. - = Not applicable; ✓ = Requirement is likely satisfied; ! = Significant difficulty; !! = Major difficulty.

Are There General Facts about Compassion and Honesty?

We assume that there are general facts about compassion and honesty, so that the realism condition is satisfied. As with general morality, the *particularist* argues that what is compassionate or honest varies situation to situation, depending on fine-grained details that are impossible to specify in advance, the *relativist* argues that what is compassionate or honest varies depending on one's culture or group, and the *skeptic* argues that there are no facts whatsoever about what is genuinely compassionate or honest.

For compassion and honesty, even more so than for general morality, the most extreme forms of skepticism, relativism and particularism are implausible: Needlessly torturing babies is not compassionate in any culture or context. Misleading those who trust and depend on you, simply for some small personal advantage and at great cost to them, could never rightly be interpreted as the pinnacle of honesty. However, as with general morality, moderate forms of relativism and particularism create conceptual challenges. The actions that constitute compassion or honesty might substantially vary between groups or depending on fine details of the situation. In a particular social context, is a white lie considered dishonest? In a large city, does not stopping to help constitute a failure of compassion? Fixed measures risk insensitivity to cultural and contextual variability in appropriate standards of honesty and compassion. Flexible measures can better avoid inappropriate rigidity, but only if employed by judges who are knowledgeable of group norms and relevant contextual details.

Although the challenges of relativism and particularism for compassion and honesty remain substantial, we judge them to be somewhat less severe than the parallel conceptual issues for general morality. Compassion and honesty are more tightly conceptually connected to

specific types of behavior than is general morality. Accordingly, there is likely to be more cross-cultural agreement and cross-situational consistency concerning what is compassionate or honest than concerning what is morally good overall. Whether it's morally bad not to stop to help a beggar might be highly culturally and situationally variable, but except in highly unusual circumstances, it is less compassionate than stopping to help. Similarly, whether it's morally bad to underreport income to the government varies situationally and culturally, but it is less honest than accurately reporting income.

How Should We Determine What is (Un)compassionate or (Dis)honest?

Idiosyncratic Understandings. As with flexible measures of general morality, disagreement on or ignorance about what constitutes compassion or honesty could lead to inaccuracies. For example, people might disagree on whether abortion is a compassionate act, depending on whether they are focusing on consequences for mother or consequences for the fetus. They might also disagree on whether failing to disclose important details (i.e., lies of omission) is dishonest. People who are uncompassionate or dishonest could also ignorantly rate themselves or their counterparts as being highly compassionate or honest.

However, it is probably reasonable to assume that few judges would describe lying and stealing as "honest" (even if morally justified) or neglecting a friend in need as "compassionate" (even if excusable given the circumstances). Idiosyncratic conceptual divergence, like divergence due to relativistic and particularist considerations, is therefore likely to be more limited for familiar, specific moral traits than for general morality.

Commonsense Understandings. Commonsense understandings of compassion and honesty might suffer from the same biases that apply to commonsense understandings of morality described above. For example, a 19th century United States citizen's understanding of

compassion might have conveniently excluded the treatment of slaves. A commonsense conception of honesty might treat the explicit utterance of falsehoods as much more dishonest than paltering in the sense of intentionally misleading people while uttering statements that are each, strictly speaking, true (Rogers et al., 2017), whereas (arguably) a more ethically correct conception of honesty might treat uttering falsehoods and paltering as similarly dishonest (Cooper et al., 2023). Nevertheless, given the more limited range of reasonable disagreement regarding the application of “compassionate” and “honest” than of “moral”, commonsense understandings are likely to be more accurate for the specific concepts of compassion and honesty than for the broader concept of morality.

Ethical Frameworks. Alternatively, researchers can employ an ethical framework to avoid the pitfalls of idiosyncratic and commonsense understandings of compassion and honesty. For example, a consequentialist might suggest that honesty is best measured by observable behaviors or by the extent to which one’s actions lead others to form accurate beliefs. A deontologist might operationalize honesty as the extent to which a person abides by a rule (e.g., to never lie). Virtue ethicists might emphasize the importance of having the right, virtuous motivation (Miller, 2017; Roberts & West, 2020; Wilson, 2018) and the optimal balance between excess and deficiency on a trait (Ng & Tay, 2020). While this characterization is somewhat simplistic, it illustrates that even in measuring the specific traits of compassion and honesty, researchers cannot avoid conceptual commitments (Wright et al., 2020).

Still, for the reasons articulated above, it remains likely that the range of reasonable expert ethical disagreement about honesty and compassion will be somewhat more limited than concerning overall general morality. For example, Kant held that one should never lie under any circumstances—even, notoriously, to a murderer at your door who is seeking a person hiding

under your protection (Kant, 1797/1996). Consequentialists, in contrast, generally hold that lying is sometimes morally permissible or even morally required, if it leads to good outcomes. Kant and the consequentialists disagree much more sharply about what constitutes an act of general morality than about what constitutes an act of (dis)honesty.

How Should We Weight Different Aspects of Morality?

Compassion and honesty, like general morality, have multiple facets, and different psychological and philosophical definitions emphasize different features. Compassion researchers vary in their emphasis on understanding and empathizing with others' emotions, recognizing the interconnectedness of human beings, deep awareness of others' suffering, caring about others' welfare, feelings of pity or concern for others' suffering, and the desire to alleviate others' suffering via altruistic behaviors (Bloom, 2017; Gilbert, 2017; Goetz et al., 2010; Nussbaum, 1996). According to Miller (2017), honesty comprises at least five different types of behaviors: An honest person is disposed to reliably tell the truth, to not steal, follow the rules in a situation when they are fair and appropriate, keep reasonable promises, and to give a complete presentation of the facts. Some conceptualizations of honesty also emphasize features of value-adherence (Peterson & Seligman, 2004), truth-seeking and intellectual honesty (Guenin, 2005) and encouraging others to form accurate beliefs (Cooper et al., 2023b).

To measure a person's general compassion or honesty, these various facets of compassion and honesty need to be commensurable into a single metric, and we must decide how to weigh these facets. Arguably, the components of compassion can be interpreted as reflecting a good-willed responsiveness to other living beings (Cokelet, 2018), and the components of honesty reflect a concern with fostering beliefs that accurately reflect reality (Miller, 2017; Wilson, 2018). However, it's unclear how various facets (e.g., truth-telling vs.

refraining from stealing), thoughts, feelings, motivations, behaviors, omissions, consequences, and conduct within different roles should be precisely weighted. For example, is a person more “compassionate” if they care deeply about others’ welfare but rarely act on these sympathetic feelings, or if they do helpful things without feeling much emotional concern (for similar debates about the definition of altruism, see Pfattheicher et al., 2022)? How diagnostic are single failures (e.g., a single data fabrication arguably constitutes absence of research honesty while a single failure of compassion might not constitute an absence of compassion; Trafimow & Trafimow, 1999)? Researchers need to make theoretical decisions about how to weigh the subcomponents of even specific moral traits, and these answers may only be approximations.

Under plausible assumptions, the approximately correct weighting of the facets of compassion and honesty should be strictly less difficult than the approximately correct weighting of overall morality. For simplicity, suppose that overall morality has four components, compassion, honesty, fairness, and purity, and suppose that each of these components has four facets. The correct weighting of overall morality would then require the correct weighting not only of the facets of each component but also of the components relative to each other (maybe purity matters much less than compassion, for example). On an act-based conception, similar considerations apply: A correct weighting of overall morality requires not only correctly weighting various honest and dishonest acts against each other, but also weighting those honest and dishonest acts against various compassionate and uncompassionate, fair and unfair, pure and impure acts. To the extent that all such weightings can only be approximate, general morality adds an extra dimension of error. Furthermore, different dishonest acts are probably closer to being commensurable with each other than are acts across different virtue types.

Does the Measure Apply Clearly and Consistently Across People and Time?

As with general measures of morality, specific measures of compassion or honesty can only succeed if they retain the same, clear meaning regardless of who is conducting the moral evaluation, and if they apply consistently across people, groups, and time. Flexible measures of compassion and honesty will be non-transparent to a substantial extent, since researchers will not know what specifically drove these evaluations. Furthermore, judges are likely to employ somewhat different criteria for compassion and honesty. Fixed measures employ consistent and transparent criteria, but risk failing to appropriately account for cultural and situational variation. Failing to email the owner of a lost wallet might seem to be the same, dishonest moral action across social groups and time (Cohn et al., 2019). But in collectivistic cultures, the more socially restrained act of safeguarding the wallet until the owner returns to retrieve it may be construed as being more relevant to personal integrity (Yang et al., 2023). Similarly, the “same” compassionate act of helping a confederate who appears to be having a medical emergency in the street might have very different perceived risks in a small town compared to a big city. Fixed measures of compassion and honesty risk presenting a superficial appearance of transparent consistency across participants while obscuring large underlying differences in the actual or psychological moral significance of the acts involved.

However, if we are correct that there is likely to be less disagreement—across cultures, groups, situations, and expert frameworks—about what constitutes compassion or honesty than about what constitutes morality more broadly construed, then these downsides are likely to be less serious for measures of compassion or honesty than for general morality. If judges are more likely to rely on similar inputs in judging honesty or compassion than in judging general morality, inconsistency will be lower for the former than the latter. Similarly, although researchers will not know specifically what (un)compassionate or (dis)honest behaviors or

mental states are influencing judges' ratings, they will have a clearer idea of the broad types of behaviors and mental states at issue than for general morality ratings, which could be driven by anything from purity violations to political activism to feelings of compassion for farm animals. Also, although the overall moral significance of failing to stop a stranger who needs help might vary considerably depending on the situational context, the fact that it is less compassionate than helping remains approximately the same. Greater definitional agreement also means that such measures are likely to retain more of their actual and psychological moral relevance when applied across people, groups, and time. It also seems less likely that researchers' fixed definitions of compassion or honesty (compared to general morality) would be unrecognizably alien to the people who are being measured.

Methodological Requirements for Measuring Specific Moral Traits

Methodologically, too, similar troubles arise for the measurement of particular moral traits, such as compassion and honesty, as arise for general morality, though often in less severe form (see Table S3).

Table S3

Methodological Requirements for Measuring General Morality, Compassion, or Honesty using Self-Report, Reputation-Based, Behavioral, or Biological Measures

Methodological requirement	General Morality				Compassion				Honesty			
	SR	Rep.	Behav.	Bio.	SR	Rep.	Behav.	Bio.	SR	Rep.	Behav.	Bio.
Judges have full information about the target's relevant characteristics.	✓	!!	-	-	✓	!	-	-	✓	!!	-	-
Judges correctly use this information to form an unbiased impression of the target's traits.	!!	!	-	-	!	!	-	-	!	!	-	-
Judges are willing to truthfully report their impressions of the target.	!!	!!	-	-	!	!	-	-	!	!	-	-
The measure represents a holistic evaluation of the target's general tendencies.	✓	✓	!!	!!	✓	✓	!!	!!	✓	✓	!!	!!
The measure is feasible enough to be implemented at scale.	✓	!	!!	!!	✓	!	!	!!	✓	!	!	!!

Note. SR = Self-report, Rep. = Reputation, Behav. = Behavior, Bio. = Biological measures. - = Not applicable; ✓ = Requirement is likely satisfied; ! = Significant difficulty; !! = Major difficulty. Reputation = assume a best-case scenario where there are multiple judges who know the target from different domains of life.

When self-reporting one's own levels of compassion and honesty, judges have access in principle to essentially the full range of relevant information, but are likely to have self-serving ego-protective biases and excessive focus on good intentions (Klein & Epley, 2017; Vazire, 2010). Informants are likely to be somewhat less biased but will have considerably less information, perhaps especially for dishonesty if dishonest targets are able to successfully hide their dishonest behavior from judges who don't know them especially well. Even when judges have accurate impressions, social desirability is likely to impact self-ratings of compassion and honesty, and resistance to reporting negative impressions of targets are likely to impact informant ratings. Rating inversions are possible, for example, if honest judges are more likely to report lapses of honesty in themselves or if highly compassionate judges are more acutely aware of lapses of compassion in their somewhat compassionate friends than uncompassionate judges are in their possibly even less compassionate friends.

Some of these problems are likely to be less severe for self- or informant-reports of compassion and honesty compared to general morality. Accurate reports of general morality requires that judges have full information about each of the constituent facets of morality, which is a more difficult requirement to meet compared to having full information about only one trait (e.g., compassion or honesty). It is less clear whether self- and other-evaluations would be less susceptible to self-enhancement biases for compassion or honesty compared to general morality. On the one hand, Dunning and colleagues (1989) find that people provide more self-serving assessments when trait terms are more ambiguous (i.e., can describe a wide variety of behaviors), thus providing more room for people to use idiosyncratic criteria when forming these judgments. This suggests that there would be less room for self-enhancement for the specific traits of honesty and compassion than for general morality (at least when general morality is flexibly

assessed using broad descriptors such as “moral,” “ethical,” or “virtuous”). On the other hand, because being moral requires having high levels of many different moral traits, people might also recognize that there is a higher bar to be able to claim that they are highly “moral” compared to being able to claim that they are highly “compassionate” or “honest.” If so, people might self-enhance to a similar or greater extent on specific moral traits. However, it does seem less socially undesirable (for self-reports) and less judgmental (for informant reports) to say that you or a friend has a particular moral vice than to say that you or a friend are overall morally bad. Empirical evidence is currently indirect and mixed, with some studies finding that people self-report higher levels of general morality compared to some specific moral traits (Helzer et al., 2014) and others finding the reverse (Furr et al., 2022; Sun & Goodwin, 2020).

Like general morality, the behavioral measurement of compassion and honesty raises serious methodological questions concerning how representative and diagnostic any particular behavioral measure could be. People may engage in apparently honest or compassionate acts for reasons that are unrelated to honesty (e.g., not cheating on an exam to avoid punishment) or compassion (e.g., being vegetarian for personal health).

It is not quite as dubious to leap from, say, three behavioral measurements of honesty or compassion to a conclusion about general honesty or general compassion as it is to leap from three behavioral measurements of three different moral traits (e.g., one act of honesty, one act of compassion, and one act of fairness) to a conclusion about general morality (as we mention in the main body of the text, at some point, the “duck test” begins to apply). Nevertheless, we hold that even for compassion and honesty such inferences are highly problematic and their difficulty should be emphasized. This is perhaps especially true for the behavioral measures that are most feasible to implement at scale, such as laboratory measures, which often lack ecological validity,

or social media interactions, which might be unrepresentative of face-to-face interactions. In line with the principle that matching predictors with criteria enhances validity (i.e., the bandwidth-fidelity tradeoff; Ashton et al., 2014; Hogan & Roberts, 1996; Ones & Viswesvaran, 1996), few-shot behavioral measures are likely better indicators of narrow traits like honesty-on-taxes or compassion-for-nearby-animals than for broad personality traits like honesty and compassion. Similar remarks apply to biological measures: Skin conductance, structural brain differences, and differences in brain activity likely remain poor measures of general compassion and honesty even if it's not quite as absurd to infer a person's level of compassion as it is to infer their general morality from fMRI activity response to pictures of suffering children.

4. Conceptual and Methodological Requirements for Measuring Non-Moral Traits

Table S4

Conceptual Requirements for Constructing Flexible vs. Fixed Measures of Extraversion (E) or Well-Being (WB)

Conceptual requirement	Flexible		Fixed	
	E	WB	E	WB
1. There must be general facts about a person's extraversion/well-being.				
Realism: There must be facts about what is extraverted vs. introverted / good or bad for well-being; <i>and</i>	✓	✓	✓	✓
Universalism: The same general things must be extraverted vs. introverted / good vs. bad for well-being for different people or in different groups; <i>and</i>	✓	!	✓	!
Generalism: What constitutes extraversion/well-being must not depend on highly particular features of specific situations.	✓	!	✓	!
2. The measure must correctly identify which characteristics are extraverted vs. introverted / good vs. bad for well-being.				
Judges' idiosyncratic judgments are correct; <i>or</i>	!	!	-	-
Commonsense understandings, as operationalized by the researchers, is correct; <i>or</i>	-	-	!	!
The favored expert framework is correct.	-	-	✓	!
3. The measure must correctly weigh different components of extraversion/well-being against each other.				
Unity: Extraversion/well-being must be a coherent construct; <i>and</i>	✓	!	✓	!
Commensurability: There must be a common "currency" in which components of extraversion/well-being can be appropriately compared.	!	!!	!	!!
4. The measure must apply clearly and consistently across people and time.				
Transparency: It must be clear what is being measured; <i>and</i>	!	!	✓	✓
Equivalence:				
A fixed measure must capture the same thoughts, feelings, and/or behaviors across people; <i>and</i>	-	-	✓	✓
A fixed measure must have the same significance for extraversion/well-being across people; <i>or</i>	-	-	✓	!
For flexible measures, judges must use the same criteria to judge a person's extraversion/well-being; <i>and, depending on the aim</i>	!	!!	-	-
The measure must be psychologically relevant.	✓	!	✓	!

Note. - = Not applicable; ✓ = Requirement is likely satisfied; ! = Significant difficulty; !! = Major difficulty.

Table S5

Methodological Requirements for Measuring Extraversion (E) or Well-Being (WB) using Self-Report, Reputation-Based, Behavioral, or Biological Measures

Methodological requirement	Extraversion				Well-Being			
	SR	Rep.	Behav.	Bio.	SR	Rep.	Behav.	Bio.
Judges have full information about the target's relevant characteristics.	✓	✓	-	-	!	!!	-	-
Judges correctly use this information to form an unbiased impression of the target's traits.	✓	✓	-	-	!	!	-	-
Judges are willing to accurately report their impressions of the target.	✓	✓	-	-	✓	✓	-	-
The measure represents a holistic evaluation of the target's general tendencies.	✓	✓	!	!!	✓	✓	!!	!!
The measure is feasible enough to be implemented at scale.	✓	✓	!	!!	✓	✓	!!	!!

Note. SR = Self-report, Rep. = Reputation, Behav. = Behavior, Bio. = Biological measures. - = Not applicable; ✓ = Requirement is adequately satisfied; ! = Significant difficulty; !! = Major difficulty. Reputation = assume a best-case scenario where there are multiple raters who know the target from different domains of life.

References

- Allegrini, A. G., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J. B., & Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. *Molecular Psychiatry*, *24*(6), Article 6. <https://doi.org/10.1038/s41380-019-0394-4>
- Ashton, M. C., Paunonen, S. V., & Lee, K. (2014). On the validity of narrow and broad personality traits: A response to Salgado, Moscoso, and Berges (2013). *Personality and Individual Differences*, *56*, 24–28. <https://doi.org/10.1016/j.paid.2013.08.019>
- Badhwar, N. K. (1996). The limited unity of virtue. *Noûs*, *30*(3), 306–329. <https://doi.org/10.2307/2216272>
- Blair, R. J. R. (2013). The neurobiology of psychopathic traits in youths. *Nature Reviews Neuroscience*, *14*(11), Article 11. <https://doi.org/10.1038/nrn3577>
- Bloom, P. (2017). Empathy and its discontents. *Trends in Cognitive Sciences*, *21*(1), 24–31. <https://doi.org/10.1016/j.tics.2016.11.004>
- Bouchard, T. J. (2004). Genetic influence on human psychological traits: A survey. *Current Directions in Psychological Science*, *13*(4), 148–151. <https://doi.org/10.1111/j.0963-7214.2004.00295.x>
- Brethel-Haurwitz, K. M., Cardinale, E. M., Vekaria, K. M., Robertson, E. L., Walitt, B., VanMeter, J. W., & Marsh, A. A. (2018). Extraordinary altruists exhibit enhanced self–other overlap in neural responses to distress. *Psychological Science*, *29*(10), 1631–1641. <https://doi.org/10.1177/0956797618779590>
- Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science*, *365*(6448), 70–73. <https://doi.org/10.1126/science.aau8712>

Cokelet, B. (2018). The virtues of compassion. In *The Moral Psychology of Compassion* (pp.

15–32). Rowman & Littlefield International. <https://philpapers.org/rec/COKTVO>

Cooper, B., Cohen, T. R., Huppert, E., Levine, E. E., & Fleeson, W. (2023). Honest behavior:

Truth-seeking, belief-speaking, and fostering understanding of the truth in others.

Academy of Management Annals, *17*(2), 655–683.

<https://doi.org/10.5465/annals.2021.0209>

Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate?: Testing the

theory of morality-as-cooperation in 60 societies. *Current Anthropology*, *60*(1), 47–69.

<https://doi.org/10.1086/701478>

Dahl, A. (2023). What we do when we define morality (and why we need to do it).

Psychological Inquiry, *34*(2), 53–79, <https://doi.org/10.1080/1047840X.2023.2248854>

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The

role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of*

Personality and Social Psychology, *57*(6), 1082–1090. [https://doi.org/10.1037/0022-](https://doi.org/10.1037/0022-3514.57.6.1082)

[3514.57.6.1082](https://doi.org/10.1037/0022-3514.57.6.1082)

Furr, M., Prentice, M., Hawkins Parham, A., & Jayawickreme, E. (2022). Development and

validation of the Moral Character Questionnaire. *Journal of Research in Personality*, *98*,

104228. <https://doi.org/10.1016/j.jrp.2022.104228>

Gilbert, P. (2017). *Compassion: Concepts, research and applications*. Taylor & Francis.

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis

and empirical review. *Psychological Bulletin*, *136*(3), 351–374.

<https://doi.org/10.1037/a0018807>

- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). *Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism* (P. Devine & A. Plant, Eds.; Vol. 47, pp. 55–130). Academic Press. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. <https://doi.org/10.1037/a0021847>
- Guenin, L. M. (2005). Intellectual honesty. *Synthese*, *145*(2), 177–232. <https://doi.org/10.1007/s11229-005-3746-3>
- 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. <https://doi.org/10.1177/0146167214554957>
- Hogan, J., & Roberts, B. W. (1996). Issues and non-issues in the fidelity–bandwidth trade-off. *Journal of Organizational Behavior*, *17*(6), 627–637. [https://doi.org/10.1002/\(SICI\)1099-1379\(199611\)17:6<627::AID-JOB2828>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1099-1379(199611)17:6<627::AID-JOB2828>3.0.CO;2-F)
- Horgan, T., & Timmons, M. (2005). Moral phenomenology and moral theory. *Philosophical Issues*, *15*, 56–77. <https://doi.org/10.1111/j.1533-6077.2005.00053.x>
- Jach, H. K., Feuerriegel, D., & Smillie, L. D. (2020). Decoding personality trait measures from resting EEG: An exploratory report. *Cortex*, *130*, 158–171. <https://doi.org/10.1016/j.cortex.2020.05.013>
- Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology*, *96*(3), 521–537. <https://doi.org/10.1037/a0013779>

- Kant, I. (1797). On a supposed right to lie from philanthropy. In M. J. Gregor (Ed.), *Immanuel Kant: Practical Philosophy*. Cambridge University Press.
- Klein, N., & Epley, N. (2017). Less evil than you: Bounded self-righteousness in character inferences, emotional reactions, and behavioral extremes. *Personality and Social Psychology Bulletin*, 43(8), 1202–1212. <https://doi.org/10.1177/0146167217711918>
- Lykken, D. T. (1978). The psychopath and the lie detector. *Psychophysiology*, 15(2), 137–142. <https://doi.org/10.1111/j.1469-8986.1978.tb01349.x>
- Macintyre, A. (1957). What morality is not. *Philosophy*, 32(123), 325–335. <https://doi.org/10.1017/S0031819100051950>
- Magnusson, K. (2023). *A causal inference perspective on therapist effects*. <https://osf.io/preprints/psyarxiv/f7mvz>
- Malle, B. F. (2021). Moral judgments. *Annual Review of Psychology*, 72(1), 293–318. <https://doi.org/10.1146/annurev-psych-072220-104358>
- Marsh, A. A., Stoycos, S. A., Brethel-Haurwitz, K. M., Robinson, P., VanMeter, J. W., & Cardinale, E. M. (2014). Neural and cognitive characteristics of extraordinary altruists. *Proceedings of the National Academy of Sciences*, 111(42), 15036–15041. <https://doi.org/10.1073/pnas.1408440111>
- McGraw, A. P., & Warren, C. (2010). Benign violations: Making immoral behavior funny. *Psychological Science*, 21(8), 1141–1149. <https://doi.org/10.1177/0956797610376073>
- Miller, C. B. (2017). Honesty. In W. Sinnott-Armstrong & C.B. Miller (Eds.), *Moral psychology: Virtue and character* (pp. 237–273). Boston Review. <https://psycnet.apa.org/record/2017-17609-018>

- Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, *6*(10), Article 10.
<https://doi.org/10.1038/nrn1768>
- Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, *125*(5), 656–688. <https://doi.org/10.1037/rev0000111>
- Ng, V., & Tay, L. (2020). Lost in translation: The construct representation of character virtues. *Perspectives on Psychological Science*, *15*(2), 309–326.
<https://doi.org/10.1177/1745691619886014>
- Nussbaum, M. (1996). Compassion: The basic social emotion. *Social Philosophy and Policy*, *13*(1), 27–58. <https://doi.org/10.1017/S0265052500001515>
- Ones, D. S., & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior*, *17*(6), 609–626.
[https://doi.org/10.1002/\(SICI\)1099-1379\(199611\)17:6<609::AID-JOB1828>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1099-1379(199611)17:6<609::AID-JOB1828>3.0.CO;2-K)
- Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. American Psychological Association.
- Pfattheicher, S., Nielsen, Y. A., & Thielmann, I. (2022). Prosocial behavior and altruism: A review of concepts and definitions. *Current Opinion in Psychology*, *44*, 124–129.
<https://doi.org/10.1016/j.copsyc.2021.08.021>
- Plomin, R., & von Stumm, S. (2022). Polygenic scores: Prediction versus explanation. *Molecular Psychiatry*, *27*(1), Article 1. <https://doi.org/10.1038/s41380-021-01348-y>
- Portmore, D. W. (2008). Are moral reasons morally overriding? *Ethical Theory and Moral Practice*, *11*(4), 369–388. <https://doi.org/10.1007/s10677-008-9110-1>

- Roberts, R. C., & West, R. (2020). The virtue of honesty: A conceptual exploration. In C. B. Miller & R. West (Eds.), *Integrity, Honesty, and Truth Seeking*. Oxford University Press.
<https://doi.org/10.1093/oso/9780190666026.003.0004>
- Rogers, T., Zeckhauser, R., Gino, F., Norton, M. I., & Schweitzer, M. E. (2017). Artful paltering: The risks and rewards of using truthful statements to mislead others. *Journal of Personality and Social Psychology*, *112*(3), 456–473.
<https://doi.org/10.1037/pspi0000081>
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, *76*(4), 574–586. <https://doi.org/10.1037/0022-3514.76.4.574>
- Russell, D. C. (2009). *Practical intelligence and the virtues*. OUP Oxford.
- Saxe, L., & Ben-Shakhar, G. (1999). Admissibility of polygraph tests: The application of scientific standards post- Daubert . *Psychology, Public Policy, and Law*, *5*(1), 203–223.
<https://doi.org/10.1037/1076-8971.5.1.203>
- Silventoinen, K. (2003). Determinants of variation in adult body height. *Journal of Biosocial Science*, *35*(2), 263–285. <https://doi.org/10.1017/S0021932003002633>
- Sinnott-Armstrong, W., & Wheatley, T. (2012). The disunity of morality and why it matters to philosophy. *The Monist*, *95*(3), 355–377. <https://doi.org/10.5840/monist201295319>
- Stocker, M. (1992). *Plural and conflicting values*. Oxford University Press.
- Sun, J., & Goodwin, G. P. (2020). Do people want to be more moral? *Psychological Science*, *31*(3), 243–257. <https://doi.org/10.1177/0956797619893078>

- Tessman, L. (2014). *Moral failure: On the impossible demands of morality*. Oxford University Press.
- Trafimow, D., & Trafimow, S. (1999). Mapping perfect and imperfect duties onto hierarchically and partially restrictive trait dimensions. *Personality and Social Psychology Bulletin*, 25(6), 687–697. <https://doi.org/10.1177/0146167299025006004>
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
- Urmson, J. O. (1958). Saints and heroes. In A. I. Melden (Ed.), *Essays in Moral Philosophy* (pp. 198–216). University of Washington Press.
- Vazire, 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. <https://doi.org/10.1037/a0017908>
- Warnock, G. J. (1971). *The object of morality*. London: Meuthen.
- Warrier, V., Toro, R., Chakrabarti, B., Børghlum, A. D., Grove, J., Hinds, D. A., Bourgeron, T., & Baron-Cohen, S. (2018). Genome-wide analyses of self-reported empathy: Correlations with autism, schizophrenia, and anorexia nervosa. *Translational Psychiatry*, 8(1), Article 1. <https://doi.org/10.1038/s41398-017-0082-6>
- Wertz, J., Caspi, A., Belsky, D. W., Beckley, A. L., Arseneault, L., Barnes, J. C., Corcoran, D. L., Hogan, S., Houts, R. M., Morgan, N., Odgers, C. L., Prinz, J. A., Sugden, K., Williams, B. S., Poulton, R., & Moffitt, T. E. (2018). Genetics and crime: Integrating new genomic discoveries into psychological research about antisocial behavior. *Psychological Science*, 29(5), 791–803. <https://doi.org/10.1177/0956797617744542>

- Williams, B. A. O. (1965). Ethical consistency. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 39, 103–124.
- Wilson, A. T. (2018). Honesty as a virtue. *Metaphilosophy*, 49(3), 262–280.
<https://doi.org/10.1111/meta.12303>
- Wright, J. C., Warren, M. T., & Snow, N. E. (2020). *Understanding Virtue: Theory and Measurement*. Oxford University Press.
- Yang, Q., Zhang, W., Liu, S., Gong, W., Han, Y., Lu, J., Jiang, D., Nie, J., Lyu, X., Liu, R., Jiao, M., Qu, C., Zhang, M., Sun, Y., Zhou, X., & Zhang, Q. (2023). Unraveling controversies over civic honesty measurement: An extended field replication in China. *Proceedings of the National Academy of Sciences*, 120(29), e2213824120.
<https://doi.org/10.1073/pnas.2213824120>
- You, C., Zhou, Z., Wen, J., Li, Y., Pang, C. H., Du, H., Wang, Z., Zhou, X.-H., King, D. A., Liu, C.-T., & Huang, J. (2021). Polygenic scores and parental predictors: An adult height study based on the United Kingdom biobank and the Framingham Heart Study. *Frontiers in Genetics*, 12, 669441. <https://doi.org/10.3389/fgene.2021.669441>