



HHS Public Access

Author manuscript

Psychol Inq. Author manuscript; available in PMC 2024 September 13.

Published in final edited form as:

Psychol Inq. 2023 ; 34(2): 53–79. doi:10.1080/1047840x.2023.2248854.

What We Do When We Define Morality (And Why We Need to Do It)

Audun Dahl

University of California, Santa Cruz

Abstract

All psychological research on morality relies on definitions of morality. Yet the various definitions often go unstated. When unstated definitions diverge, theoretical disagreements become intractable, as theories that purport to explain “morality” actually talk about very different things. This article argues for the importance of defining morality and considers four common ways of doing so: The *linguistic*, the *functionalist*, the *evaluating*, and the *normative*. Each has encountered difficulties. To surmount those difficulties, I propose a *technical*, *psychological*, *empirical*, and *distinctive* definition of morality: obligatory concerns with others’ welfare, rights, fairness, and justice, as well as the reasoning, judgment, emotions, and actions that spring from those concerns. By articulating workable definitions of morality, psychologists can communicate more clearly across paradigms, separate definitional from empirical disagreements, and jointly advance the field of moral psychology.

Keywords

moral psychology; moral philosophy; definitions; meta-theory

“I’ll tell you what real love is,” Mel said. “I mean, I’ll give you a good example. And then you can draw your own conclusions.” He poured more gin into his glass. He added an ice cube and a sliver of lime. We waited and sipped our drinks. Laura and I touched knees again. I put a hand on her warm thigh and left it there.

“What do any of us really know about love?” Mel said.

Raymond Carver (1981), “What we talk about when we talk about love,” p. 176

In Raymond Carver’s (1981) short story “What we talk about when we talk about love,” two couples hold a conversation about love that grows increasingly tense and drunken. The four protagonists try to talk about the same thing but lack a shared notion of love. The last word goes to the loudest of the party—the cardiologist Mel—before the conversation awkwardly dies. For anyone who has studied moral psychology, it is easy to imagine a similar story about four psychologists discussing morality. In the latter story, each psychologist would make claims based on their own, unstated definition of morality. Each protagonist, and every

reader of the story, would be unsure what the other three were talking about when they said “morality.” This, I will argue, is the predicament of today’s science of moral psychology.

Seeking a way out of this predicament, I will engage the four protagonists, each representing a contemporary approach to defining morality (see e.g., Machery, 2018; Stich, 2018). One protagonist defines morality as whatever people call “moral” (the *linguistic* approach); a second defines morality in terms of its social function (the *functionalist*); a third defines morality as the collection of right actions (the *evaluating*); and a fourth defines morality to comprise all judgments about right and wrong (the *normative*). These four ways of defining morality are not themselves theories of morality, although they undergird such theories. They are different ways of demarking the domain that psychological theories of morality seek to explain, just as definitions of planets demark the domain of planetary science.

In this article, I will first argue that it is both possible and useful for psychologists to define morality, countering the arguments of some scholars (e.g., Greene, 2007; Stich, 2018; Wynn & Bloom, 2014). Without definitions, disagreements about morality will stalemate; with definitions, we can research and communicate more clearly. I will then discuss how, despite their intuitive appeal, each of the four approaches to defining morality—the linguistic, functionalist, evaluating, and normative—runs into difficulties.

The last part of the article will seek a definition of morality that achieves its goals without the corresponding difficulties. In brief, I propose to define morality as obligatory concerns with others’ welfare, rights, fairness, and justice, as well as the reasoning, judgment, emotions, and actions that spring from those concerns (Dahl et al., 2018; Dahl & Killen, 2018a; Turiel, 1983; Turiel & Dahl, 2019).¹ This definition is *technical* in that it achieves considerable, but not complete, overlap with ordinary language use; it is *psychological* in that it captures psychological characteristics that most people in all communities have; it is *empirical* in that it does not rely on researchers’ ideas about morally good and bad behaviors; and it is *distinctive* in that it distinguishes moral judgments from other normative judgments. I will not claim that the proposed definition is the only technical, psychological, empirical, and distinctive definition of morality we might adopt. I claim only that the proposed definition avoids some common problems and has proven useful for empirical research.

Do We Need to Define Morality to Study it?

Two transformational decades of psychological research on morality have intensified the need for definitions (Kohlberg, 1971; Turiel, 1983, 2010). The 2000s saw seminal publications on moral reasoning, judgments, emotions, and actions across the lifespan and in diverse cultural groups (e.g., Blake et al., 2015; Bloom, 2013; Greene et al., 2001; Haidt, 2001; Haidt & Graham, 2007; Hamlin et al., 2007). In the intervening years, scholars built on these advances in thousands of publications (see Cohen Priva & Austerweil, 2015;

¹Though I build on the arguments of Social Domain Theory (Killen & Smetana, 2015; Royzman et al., 2009; Turiel, 1983, 2015), my arguments differ in at least two key respects. First, Turiel’s (1983) original arguments have been criticized for presuming too much agreement among philosophers about the definition of morality (Machery & Stich, 2022). My proposals about how to define morality presume no such philosophical consensus. Second, my proposed definition goes beyond Turiel’s original formulations by explicitly including emotions and actions (Turiel & Dahl, 2019).

Ellemers et al., 2019; Gray & Graham, 2018; Malle, 2021; Thompson, 2012). Adding to the empirical wealth, the vast psychological literature on morality overlaps—often seamlessly—with other psychological literatures on prosociality and social norms (e.g., Legros & Cislighi, 2020; Padilla-Walker & Carlo, 2016).

For all its empirical advances, psychological research on morality has lacked theoretical integration. Some scholars have argued that morality is innate, while others have argued that morality first emerges around age three (Bloom, 2013; Carpendale et al., 2013; Hamlin, 2013; Tomasello, 2018). Many developmental psychologists have proposed that moral judgments spring from reasoning; many cognitive and social psychologists have countered that most moral judgments spring from affective intuitions and not reasoning (e.g., Greene, 2013; Haidt, 2012; Turiel, 2018). And some scholars have proposed that each cultural-historical community has its own morality, challenging the idea of a universal moral sense (Haidt & Graham, 2007; Machery, 2018; Shweder et al., 1997).

A major impediment to integration has been that moral psychologists have disagreed sharply about whether and how to define morality (Malle, 2021; Sinnott-Armstrong & Wheatley, 2012). Those who claim that morality is innate rely on different definitions of morality than those who claim that morality emerges by age three, even though the differences between their definitions mostly go unacknowledged. The same holds true for debates about the roles of emotion in morality and for debates about cultural variability. So when two psychological theories make opposing claims about “morality,” it is hard to know if they are speaking about the same thing.

The hope behind this paper is that clarity on definitions will help our field capitalize on its empirical advances and achieve greater theoretical integration. This does not mean that all scientific disagreements about morality come down to differing definitions: To know when morality emerges, we still need empirical research about what children can and cannot do at a given age. But we cannot resolve any empirical disagreements about morality without knowing what the protagonists of those disagreements mean when they write “morality.”

The Problems of Lacking Definitions

To see what happens when we lack definitions, consider the debate about whether obedience to authorities is a moral concern (see e.g., Haidt & Graham, 2007; Turiel, 2015). Is it a moral matter to follow the commands of a god, parent, or military commander? Or does that fall outside the moral and into a non-moral domain, such as a domain of social conventions? The answer depends on what the questioner means by “moral.” If “moral” means “relevant to judgments of right and wrong,” virtually all theorists would say “yes” (Haidt & Graham, 2007; Smetana, 2013; Turiel, 2015). If “moral” means “whatever people call ‘moral’” the answer will vary from person to person, and occasion to occasion, since people use “moral” in a variety of ways—and some languages have no equivalent term (Simpson & Weiner, 1989; Wierzbicka, 2007). If “moral” means “concerns with welfare, rights, justice, and fairness” (Turiel, 1983), concerns with authority commands would fall outside the moral domain. And if, as they often do, two debaters unknowingly operate with different definitions of “morality,” the debate verges on the meaningless.

Similar confusions have arisen in debates about whether all moral judgments derive from perceived harm. In their Theory of Dyadic Morality, Gray and colleagues (2012) proposed that “all moral transgressions are fundamentally understood as agency plus experienced suffering—that is, interpersonal harm” (p. 101). According to this theory, there are no “harmless wrongs” (Gray et al., 2014), and perceptions of harm prompt judgments of wrongness.

Critics of Gray’s Theory of Dyadic Morality have countered with examples that, in their view, separate morality from harm: People can judge harmless actions as immoral and harmful actions as moral (Graham & Iyer, 2012; Royzman & Borislow, 2022). In response to such critics, Gray has broadened the definition of harm to include variety consequences (e.g., Schein & Gray, 2018b). The harm in question is not just bodily harm, but virtually any actual or potential negative consequence—to a soul, a group, a society, an animal, the environment, or even oneself. With this broader notion of harm, the Theory of Dyadic Morality can defend itself against examples of moral wrongs that, while harmless in a narrow sense, are harmful in the broader sense.

But ambiguity still surrounds the word “moral.” Let us say a critic wants to challenge the Theory of Dyadic Morality by identifying harmless moral violations. Such a counterexample would have to belong to the intersection of two sets: the set of harmless actions and the set of moral violations. Schein and Gray’s (2018b) redefinition of harm sharpens the boundary of the set of harmless actions. But we also need to sharpen the boundary of the set of moral violations. To know what would count as a counter-example, we need to know what counts as a “moral” example. We need a definition of morality.

In some places, Gray and colleagues have written as if all judgments about right and wrong are “moral” judgments (Gray et al., 2014; Schein & Gray, 2018a). In a chapter entitled: “Moralization: How acts become wrong,” Schein and Gray’s (2018a) begin by asking “What makes an act wrong?” (p. 363). The answer, they suggest, “lie with perceptions of harm.” In other places, the theory distinguishes moral judgments from other judgments of right and wrong, as when Gray et al. (2012) wrote that the “presence or absence of harm also distinguishes moral transgressions from conventional transgressions” (p. 109; also Schein & Gray, 2018b, p. 36). This leaves us with two potential definitions of morality in the Theory of Dyadic Morality. Moral judgments are either all of normative judgments or a specific subset of normative judgments.

The choice between these two definitions determines the kind of evidence needed to evaluate the Theory of Dyadic Morality. The claim that *all* right-wrong judgments spring from perceptions of harm is a much broader claim, and requires more evidence, than the claim that some (moral) *subset* of right-wrong judgments springs from perceptions of harm. Moreover, if the theory took the latter path, it would need to identify that subset of normative judgments without reference to perceived harm. Otherwise, the core claims of the theory become tautological: If, say, the theory defined moral violations as “violations perceived as harmful,” it would have ruled out the possibility of “harmless” moral violations a priori (Smedslund, 1991).

Resolving the debates about the Theory of Dyadic Morality requires a definition of morality. Equipped with a definition of morality, we can test the theory's empirical hypothesis that all moral judgments are rooted in perceptions of dyadic harm. Here too, empirical claims and questions are conditional on a definition of morality.

Why Defining Morality is so Hard

All scientific inquiries depend on definitions. An inquiry begins with a question. How many planets are there? When do moral judgments emerge? Are all moral judgments based on perceptions of harm? The answer to each empirical question depends on the meaning of its words. Astronomers who want to count the number of planets in our solar system first need to decide what counts as a planet (Metzger et al., 2022). Definition in hand, the astronomer can then peer through their telescopes and decide whether Pluto, the Moon, and other heavenly bodies count as planets. (It is not strictly necessary that all astronomers use the same definition of planets, as long as each astronomer is clear on what they mean when they count their planets.) Before discovering that the structure of DNA had a double-helix structure, biologists first had to define and identify DNA (Wade, 2004). And so on.

Among scientific concepts, morality has proven unusually vexing to define. One cause for vexation is the myriad of meanings that “moral” and “morality” have in ordinary language (Graham et al., 2011; Malle, 2021). In the Oxford English Dictionary, “moral” has over twenty meanings, ranging from considerations about right, wrong, good, or bad to descriptions of practical certainty (“moral certainty,” Simpson & Weiner, 1989). Fortunately, everyday polysemy is not scientifically fatal. The many meanings of the word “cell” did not stop biologists from arriving at a workable definition—one that differentiates bodily cells from battery cells and prison cells.

A related complication is that research participants, whose morality psychologists study, have their own uses of “moral” and “morality.” Participants’ word usage do not generally align with psychologists’ word usage. (Biological cells, by contrast, do not mind their homonymy with battery cells.) To avoid such misalignment in word use, some researchers have proposed to rely on what “ordinary people” call moral (Greene, 2007). As we will see later, this solution—despite its appeal—brings its own trouble.

What is worse, “moral” can take on both evaluative and descriptive meanings (Gert & Gert, 2017; Machery & Stich, 2022). In its evaluative sense, “moral” means “morally good/right” and contrasts with “immoral.” A person may call the execution of a murderer a “moral action” in the evaluative sense if they positively *evaluate* the use of capital punishment. In its descriptive sense, “moral” means “evoking views about what is morally good or bad” and contrasts with “non-moral” or “amoral.” A person may call legal executions a “moral issue” in a descriptive sense if they mean to describe capital punishment as a topic about which people hold strong views about the rights and wrongs of murderers. “Moral” also has a descriptive meaning in the phrase “moral psychologists.” We label someone a moral psychologist because they study moral psychology—not because they are morally better than other psychologists (see Schwitzgebel & Rust, 2014). When researchers call a phenomenon “moral,” a reader may wonder whether they use “moral” in an evaluative or descriptive sense.

Such difficulties have led to pessimism about definitions. Many psychologists and philosophers have argued that defining “morality” is either unfeasible or counterproductive in psychological research (Greene, 2007; Stich, 2018; Wynn & Bloom, 2014). Perhaps sharing the same pessimism, other psychologists have tacitly avoided defining morality in their writings. Let us consider arguments against defining morality.

Arguments Against Defining Morality

One common argument is that we cannot define morality until philosophers and psychologists have reached a consensus on how to define it. Wynn and Bloom (2014) wrote that “starting with a definition is ill advised. After all, there is no agreed-upon definition of morality by moral philosophers [...], and psychologists sharply disagree about what is and is not moral” (p. 436). They argue that only a definition of morality widely shared by philosophers or psychologists would do; until we reach that agreement, morality is better left undefined.

True, philosophers and psychologists have no consensual definition of morality (Gert & Gert, 2017; Stich, 2018). But the lack of a shared definition is not usually a reason *against* explaining what we mean. More often, the lack of a shared definition is a reason for providing one’s own definition, to help readers understand us. Clear communication demands specification, especially when our words has several possible meanings. If I am flying from Paris to the London that lies in the Canadian province of Ontario, I should say which London I mean. If I am flying from Paris to the New York City that lies in the U.S. state of New York, such specification is superfluous. And if I am proposing to explain morality, I should say which morality I mean.

The definition we provide does not need to be the definitive definition. There is no shame in settling for an inquiry-specific or technical definition. Philosophers—experts in defining terms—routinely provide definitions that their colleagues do not share. In his classic book about conventions, Lewis (1969) wrote: “If the reader disagrees [with his definition of conventions], I can only remind him that I did not undertake to analyze anyone’s concept of convention but mine” (p. 46). Even if his definition differs from the definitions of others, Lewis views his philosophical investigation of conventions is a worthwhile one. For another philosophical project, another definition of conventions might be better. Nutritionist and botanists have different definitions of fruits. For nutritionists, who study healthy eating, tomatoes are vegetables; for botanists, who study the biology of plants, tomatoes are fruits (Petruzzello, 2016). Each discipline lives happily with its own definition, tailored to its purpose, and with the knowledge that others define fruits differently. These are reassuring precedents for moral psychologists. We can safely adopt a definition of morality in our own work without requiring that our definition be shared by everyone else.

Others have contended psychologists constrict their vision if they define morality. Greene (2007) wrote that “[f]or empiricists, rigorously defining morality is a distant goal, not a prerequisite. If anything, I believe that defining morality at this point is more of a hindrance than a help, as it may artificially narrow the scope of inquiry” (p. 1). Greene’s worry is that, if we define morality, we might miss out on important phenomena by defining them out of our inquiry. But if having a definition might render the inquiry too narrow, lacking

a definition might render the inquiry too broad. For without definitions, we cannot separate the phenomena that our theory seeks to explain from the phenomena that the theory does not seek to explain. When we develop a theory of morality, do we hope to explain how people judge violations of dress codes, of drink recipes, or of the manual that came with our new coffee machine? Our answers to these questions reveal the perimeter of our definition and shape the choice of stimuli for our studies.

Furthermore, the scope of our inquiry is not chained to a single, defined term. A researcher who defines morality can still choose to study things other than morality. Researchers in the Social Domain Tradition have used a fairly narrow definition of morality (issues of welfare, rights, justice, and fairness), but this has not prevented them from studying other normative issues, such as social conventions (Killen & Smetana, 2015; Turiel, 2015).

The sheer formidability of the task might be a further reason why psychologists leave morality undefined (Stanford, 2018; Stich, 2018). Spanning several millennia, debates about the nature of morality have enlisted enough intellectual giants to intimidate us into silence. Who are we to enter a fray that already has Plato and Kant in it? Luckily, we do not need to shoot for the ultimate definition of the One True Morality, forever immune to revision. Any researcher is free to abandon their old definitions if new ones turn out more useful—and scientists often do (Kuhn, 1962; Lazarus, 2001; Mukherjee, 2016). We need a car today even if we can buy a safer and more self-driving car next year. And we need a definition today even if we will find a better definition in the future.

If a reader grants that a definition of morality would be useful—even a definition not shared by all psychologists—one empirical challenge remains. Communities seem to differ radically in their values, in what their members judge as right and wrong,² and in their way of life (Henrich et al., 2010; Rogoff, 2003; Shweder et al., 1987; Turiel, 2002). Eating beef is wrong for Hindus but not for Muslims (Srinivasan et al., 2019). Brits must drive on the left side of the road, and Americans must drive on the right. In the United States, most Democrats say that abortion should be legal under most or all circumstances, whereas only a minority of Republicans agree (Gallup Inc., 2018). Faced with such cultural variability, can a single definition of morality really work for all communities? On closer inspection, the cultural differences in values and normative judgments coexist with profound cultural similarities. These cultural similarities make a universal definition of morality possible.

The Cultural Similarities that Make a Universal Definition of Morality Possible

In cross-cultural research, cultural differences usually get more attention than cultural similarities (Schwartz & Bardi, 2001; Turiel, 2002). Psychology has a history of characterizing some countries or continents as individualist and others as collectivist; some cultures as “loose” in their application of rules and others as tight (Gelfand, 2018; Triandis, 2001; for discussion, see Rogoff, 2003).

²As I explain later, I am not proposing that all matters of right and wrong should be called moral issues. For now, it is sufficient to note that people in different places differ in their judgments about right and wrong ways of acting. So that my conclusions in this section about the empirical literature would not hinge on my own definition of morality, I framed it more broadly, in terms of concerns and judgments about right and wrong.

Our attention to cultural differences tempts us to overlook the similarities in the values, principles, and judgments of most people in all communities. Around the globe, children, adolescents, and adults think it is generally wrong to cause harm, disobey legitimate authorities or traditions, or engage in dangerous behaviors (see Turiel, 2002). Several research paradigms have revealed that communities share fundamental values and norms, even if they apply those values and norms differently to different situations.

One questionnaire that has revealed such similarities is the Moral Foundations Questionnaire (MFQ). The MFQ assesses five moral “foundations” or values: Harm, fairness, group loyalty, authority, and purity (Graham et al., 2011, 2013; Haidt & Graham, 2007). The questionnaire asks participants which considerations are relevant when they “decide whether something is right or wrong.” Participants report the relevance of 15 normative considerations—three for each foundation—such as whether “someone suffered emotionally” (harm) and whether someone “showed a lack of loyalty” (loyalty). They indicate the relevance of each consideration on a 6-point scale from 1 (“not at all relevant”) to 6 (“extremely relevant”).

The original goal of the MFQ was to “gauge individual differences in the range of concerns that people consider morally relevant” (Graham et al., 2011, p. 369). Researchers have also used the MFQ in comparisons of U.S. conservatives to U.S. liberals and other cultural comparisons (Haidt & Graham, 2007). In the United States and Turkey, politically conservative participants rated concerns with authority, loyalty, and purity as more important than politically liberal participants (Haidt & Graham, 2007; Yilmaz et al., 2016). In a cross-country comparison, Zhang and Li (2015) found that participants in China, on average, rated group loyalty, authority, and purity as more important than participants in the United States. The findings with the MFQ echo prior findings by Shweder and colleagues (1987, 1997; see also J. G. Miller et al., 2017). The latter work reported that children and adults in India placed, on average, more emphasis on concerns with community and religion, and less on individual autonomy, than children and adults in the United States.

But group differences in value ratings, while statistically significant, are smaller than we think: Graham and colleagues (2012) found that U.S. adults overestimate the size of political differences in value ratings. People in all groups are concerned with harm, fairness, loyalty, authority, and purity, and the studies with the MFQ show as much. Yes, liberals give lower relevance ratings to loyalty, authority, and purity than conservatives do. But even the most extreme liberals studied by Haidt and Graham (2007) rated loyalty, authority, and purity as being, on average, between 3 (“slightly relevant”) and 4 (“somewhat relevant”) on the 6-point scale. The most extreme conservatives gave average ratings between 4 and 5 (“very relevant”). To summarize: We all care about loyalty, authority, and purity—as well as harm and fairness.

Moreover, the average cultural differences in value ratings depend on the issue people are evaluating. Conservative participants give greater weight to the commands of military officers; liberals, by contrast, place greater weight than conservatives on other authorities, such as environmentalists experts or U.S. presidents from the Democratic party (Frimer et al., 2014; Turiel, 2015). These findings show that we have to be careful when we say that

conservatives care more about authorities, loyalty, and purity in general. People are, after all, never making judgments based on authorities, group loyalties, or purities in general but always with respect to some specific authority, group, or purity.

Other research paradigms have yielded similar findings: People care about the same fundamental values everywhere, even if their weighting and specific implementation of those values vary. And even the weighting of values shows consistency. In one cross-cultural study with participants from over 50 countries, benevolence and universalism consistently emerged among the most important values (Schwartz & Bardi, 2001).

Humanity's shared basic values and concerns do not guarantee agreement on all issues. In war, the loyalties and authorities of one side conflict with the loyalties and authorities of the other. People who care about fairness can nonetheless disagree about whether the current distribution of wealth is fair (Arsenio, 2015; Jost & Kay, 2010). Still, the cross-cultural similarities in basic values and concerns provide a basis for a universally applicable definition of morality. Defining morality for psychological science is not only desirable; it is also attainable.

What We Do When We Define Morality: Four Contemporary Approaches

Once we resolve to define morality, we have to decide how to go about it. Carver's short story asks what we do when we talk about love. This section will ask what we do when we define morality. To define morality has meant different things to different psychologists. Taking divergent paths, those psychologists arrived at disparate definitions. If we reflect on what we do when we define morality in psychology, we realize that our definitions serve different ends. Some definitions serve those ends well, other definitions serve them less well. By surveying those ends, and our success in reaching them, we can make more informed choices about what to do when we define morality.

I will consider four contemporary approaches to defining morality: *the linguistic*, *the functionalist*, *the evaluating*, and *the normative* (Table 1). Each approach represents a different idea of what we do when we define morality in psychological research; each runs into its own difficulties. The four approaches to defining morality for psychological research have had multiple instantiations over the past century. Though I provide examples below, my goal is not to list all the definitions of morality proposed by psychologists, as extant papers provide excellent surveys (Graham et al., 2013; Killen & Smetana, 2015; Machery, 2018; Malle, 2021; Sinnott-Armstrong & Wheatley, 2012; Stich, 2018). My goal is to initiate a taxonomy of common and appealing ways of defining morality in psychology.

The discussion of the linguistic, functionalist, evaluating, and normative approaches will set the stage for an alternative approach to defining morality, one designed to avoid difficulties encountered by the four other approaches. I will instantiate this alternative approach by proposing a definition of morality. The definition I propose is not the only valid or useful definition of morality; it is merely *one* useful definition, better than some, possibly worse than others, and—like all definitions—a candidate for future revision.

(a) The Linguistic: Capturing How People use the Word “Moral”

One common way of defining morality in psychology, which we have already encountered, is to rely on what people call “moral” (Greene, 2007; Levine et al., 2021; Sinnott-Armstrong & Wheatley, 2012; Stich, 2018). I will call this the *linguistic* approach. Embracing a linguistic approach, Greene (2007) writes: “I and like-minded scientists choose to study decisions that ordinary people regard as involving moral judgment” (p. 1).³ Similarly, Sinnott-Armstrong and Wheatley (2012) write: “We define moral judgments to include all judgments that are intended to resemble paradigm moral judgments” (p. 356). In this approach, moral psychologists are to study anything that fits within non-scientists’ conceptions of “morality.”

Some have asserted that the linguistic approach frees psychologists from having to define morality. In their review of research on moral conviction, Skitka and colleagues (2021) write:

Rather than start with a definition of what counts as a moral concern, researchers working on moral conviction have instead asked people whether they see their position on given issues as a reflection of their personal moral beliefs and convictions. In other words, unlike most approaches that define a priori what counts as part of the moral domain, the moral conviction approach allows participants to define the degree to which their thoughts, feelings, and beliefs reflect something moral (p. 349).

But by counting as moral anything that people call “moral,” the linguistic approach is undoubtedly an approach to defining morality—even if the researchers do not put forth their own substantive definition. Via the responses of research participants, the linguistic approach offers a criterion for separating the moral from the non-moral. This separation sets the science in motion, for what falls inside the moral domain demands an explanation from moral psychologists. The linguistic approach has to hope that people’s use of the word “moral” picks out phenomena of sufficient unity to make a psychological theory possible (Sinnott-Armstrong & Wheatley, 2012).

The problem for the *linguistic* approach is that the meaning of “morality” varies enormously, both within and between persons. The same person can use the word “morality” in multiple ways: I already mentioned the twenty-plus meanings listed in the Oxford English Dictionary. All of these meanings are available to any one English speaker. This is essentially what Wittgenstein (1953) pointed out, using the example of games. In Wittgenstein’s view, we have no reason to think that any one characteristic unites board games, card games, and Olympic games—not to mention wild game (Daston, 2022). Nor have we any reason to think that any one feature unites bodily cells, battery cells, prison cells, terrorist cells, and whatever else people call “cell.” The collections of entities called “game,” “cell,” or “moral”

³I do agree with Greene (2007) that to examine everyday word use might be useful when you found a new science: “Biologists got their start, not by rigorously defining ‘life,’ but by studying the kinds of things that ordinary people regard as living” (p. 1). Without knowing exactly how the first biologists of Antiquity got going, I find Greene’s account plausible enough. When Sharp (1896) began his initial studies of moral psychology late in the 19th century, he might have benefitted from surveying what people call “morality.” Today, things are different. After over a century of data collection, the psychology of morality—by any definition—has outgrown its infancy, is no longer starting out, and is no longer a new science.

are too heterogenous and unbounded to support interesting theories. This is why biologists do not take a linguistic approach to definitions. No biologist would commit themselves to theorizing about whatever it is that people call “cells.” Biologists theorize about cells as defined by biologists and, thanks to their technical definition, can demonstrate cellular truths like: “Cells break down sugar molecules to generate energy.”

A further complication is that people differ in their use of the word “moral.” Even two people who agree completely on what is right and what is wrong might use the word “moral” in distinctly idiosyncratic ways. The Catholic Church talks about “moral certainty” and the Islamic Republic of Iran has a “morality police.” Philosophers and laypersons also use the word “morality” in disparate ways (Graham et al., 2011; Levine et al., 2021; Machery, 2018; Stich, 2018). When Malle (2021) asked 30 people to define “morally wrong,” he got 30 answers. One person answered that “morally wrong” meant “against the will of God.” Another answered: “Deliberately causing real harm is definitely a confident morally wrong. Accidental harm is more gray, and causing no one harm isn’t morally wrong.” (Also, I doubt that people’s definitions captured every way they actually use the word “moral” or “morality” in their speech.) Interpersonal variability in word use is troublesome enough when working within one language community. It borders on intractable when we look across multiple languages and time periods (Daston, 2022; Machery, 2018; Wierzbicka, 2007).

What reason do we have for assuming that, underneath these differences, everyone has a shared notion of “morality”? And, if two people disagree about what is “moral,” which persons should moral psychologists listen to when they decide what to study? Is it enough that one person calls an issue “moral” for it to count, even if that one person is being careless or manipulative in their choice of words? The upshot: Everyday language is an unreliable guide for psychologists wanting to study morality. It seems impossible for psychologists and philosophers to concoct a definition that covers all the many uses of the word “moral” (Sinnott-Armstrong & Wheatley, 2012; Wittgenstein, 1953). And even if psychologists somehow devised a definition that captured every use of the word “moral”—perhaps an extremely long and ever-changing catalog of word usage—the definition would be useless for psychological science. The phenomena falling under the definition would lack the psychological unity necessary for theorizing and hypothesis testing (Sinnott-Armstrong & Wheatley, 2012).

Still, we should not completely abandon everyday language. A *partial* overlap between the scientific and ordinary usages of “morality” seems desirable. *Some* of the things that scientists call “moral” should also be things that other people call “moral.” It would cause needless confusion to define “morality” as, say, a component of low-level color perception. A definition for scientific purposes that partially overlaps with everyday word use, but that has a singular and well-defined meaning, is a technical definition.

Technical definitions diverge from everyday word use for a reason. In scientific discourse, clear and consistent word use is essential for progress. Technical definitions ensure that readers know which phenomena our theories and hypotheses are about. In everyday language, polysemy is generally tolerable, probably unavoidable, and sometimes desirable

(Vicente & Falkum, 2017). We deploy the multiple meanings of words to make puns, obfuscate when we want to obfuscate, and insinuate when we want to insinuate. At its best, science seeks none of these. In science, polysemy produces imprecision.

Science has a history of technical definitions that partially overlaps everyday usage. From Antiquity until the Scientific Revolution of the 16th and 17th century, most people thought whales were fish (Romero, 2012).⁴ As long as “fish” meant “anything swimming in water,” whales fit the definition. Eventually, modern scientists like Carl Linnaeus (1707–1778) stopped counting whales as fish. They realized that whales—unlike most swimming creatures—are mammals who breathe with lungs and produce milk. Non-scientists, though, persisted in pooling whales and fishes. In *Moby Dick*, Herman Melville (1851)—the most famous whale aficionado of the 19th century—had the narrator Ishmael define the whale as “a spouting fish with a horizontal tail” (p. 111). Ishmael concedes that Linnaeus separated whales from fish, only to deem Linnaeus’s reasons “altogether insufficient” (p. 111). The first edition of the *Oxford English Dictionary*, published around 1900, said that in “popular language,” a fish is “any animal living exclusively in the water” (Murray & Bradley, 1900, p. 254).⁵ It added: “In modern scientific language (to which popular usage now tends to approximate),” the word “fish” is “restricted to a class of vertebrate animals provided with gills throughout life, and cold-blooded” (p. 254).

The separation of whales and fishes is an example of how scientists let technical usage overlap partially with ordinary usage (Figure 1). Even after Linnaeus put the whales with the mammals, most creatures that ordinary people called “fish”—shark, swordfish, herring—continued to meet the scientific definition of “fish.” But it was scientifically convenient to split the scientific notion of fish from certain creatures that ordinary people called fish, like whales and shellfish. It was also convenient to distinguish the scientific notion of fish from various other uses of “fish.” According to the Oxford English Dictionary, “fish” can also mean “a turtle,” “a dollar,” “a torpedo,” and—“unceremoniously”—a person, as in: “He was an odd fish” (Simpson & Weiner, 1989). Biologists eventually convinced people to stop calling whales “fish.” Yet this history began, and continues, with partial overlap between scientific and ordinary usage.

Following the example of marine biologists, moral psychologists could be looking for a *technical* definition of morality: one that sacrifices a complete overlap with everyday usage for scientific clarity. Indeed, many philosophers rely on technical definitions of morality. In doing so, they are not demanding that their technical usage capture or replace ordinary usage. In explaining his definition of “moral” values, Scanlon (1998) writes that “what seems to me most important is to recognize the distinctness of the various values I have discussed [...] Once the nature and motivational basis of these values is recognized, it does not matter greatly how broadly or narrowly the label ‘moral’ is applied” (Scanlon, 1998, p. 176; for similar arguments, see Dahl, 2014; D. Lewis, 1969). Scanlon’s main goal here is not to force others to use his definition: he does not seek to monopolize the term “morality.”

⁴I thank a reviewer for suggesting that I discuss the scientific and non-scientific usage of the word “fish.”

⁵To my surprise, the same descriptions of popular and scientific usage of “fish” remains in the contemporary edition of the Oxford English Dictionary (Simpson & Weiner, 1989). I have not met anyone who thinks whales are fishes.

His goal—one that all scholars have—was to communicate his views clearly, and a technical definition of morality did the trick.

How much overlap with everyday usage to seek—how large the gray section in Figure 1 should be—depends on at least two things. First, how much overlap to seek depends on the unity of the entities picked out by ordinary word use. If more of the phenomena identified by the ordinary usage shares basic features, they allow more theorizing. Second, it on the scientific aims behind the technical definition. Scientific theorizing about fish became easier, and richer, when theories of fish no longer had to account for whales (Sinnott-Armstrong & Wheatley, 2012, provide an analogous example from mineralogy). For psychologists, theories about morality become more interesting if the theories can explain judgments about right and wrong without also having to explain the legal notion of “moral certainty.”

No easy formula tells us where to draw the line between the technically moral and the technically non-moral. In drawing this line, we trade overlap with everyday word usage against the psychological unity of the phenomena that fall under our definition of morality. As we increase the overlap with everyday usage, we incorporate a larger and more heterogeneous set of phenomena that need explaining. Later in this article, I will discuss my proposal for where to draw the line.

(b) The Functionalist Approach: Capturing Morality by its Societal or Evolutionary Function

A second, popular approach is to define morality by its societal or evolutionary function. Exemplifying this functionalist approach, Haidt (2008) defined morality as follows: “Moral systems are interlocking sets of values, practices, institutions, and evolved psychological mechanisms that work together to suppress or regulate selfishness and make social life possible” (Haidt, 2008, p. 70; Greene, 2014; Haidt & Kesebir, 2010; Krebs, 2008; Curry et al., 2021). In a similar vein, from an evolutionary perspective, Krebs (2008) wrote: “The domain of morality pertains to the formal and informal rules and sanctions that uphold the systems of cooperation that enable members of groups to survive, to reproduce, and to propagate their genes” (p. 108). Focusing on the social function of morality, Kochanska and Aksan (2006) define morality as an inner guiding system that ensures “people’s compliance with shared rules and standards” (p. 1588). Each quote defines morality by its function or consequences, whether socially or evolutionarily. The definitions reference no psychological characteristics, such as cognitive or affective components.

When we define “morality” by its function or consequences, the risk is that we will capture phenomena that lack any shared psychological characteristics. Without shared psychological characteristics, we cannot form a meaningful psychological theory. Any number of characteristics might suppress selfishness and make social life possible, such as language, fear of punishment, and the ability to remember threats from others. Grammatical rules, for instance, make it possible to form verbal agreements of cooperation. Nonetheless, our understanding of grammatical rules have little in common with our capacity to form judgments about violence. To understand language, we will want a theory of language, not a theory of morality. Thus, even if the characteristics picked out by a functionalist theory are research-worthy, their heterogeneity impedes interesting generalizations.

There is an even bigger challenge for functionalist theories. If morality is defined by its consequences, we might not know whether a psychological characteristic belongs to the moral domain until much later. It may take minutes, days, or years to know which characteristics ultimately promote cooperation or suppress selfishness in the long run. A well-intentioned effort to help another person might strengthen cooperation in one context, whereas an identical initiative might be rebuffed as a bothersome act of interference in another context (Oakley et al., 2011). Protests against social injustice can promote peaceful coexistence in some circumstances but increase societal conflict in other circumstances (Killen, 2016; Killen & Dahl, 2020). And who would know, prior to years of systematic research, whether a constitutional right to smoke weed or own a gun would be beneficial or detrimental to “peaceful coexistence”? If we rely on the functionalist criterion, we might only know if the acts of helping or protest counted as moral long after the acts occurred, if ever. Whereas this indeterminacy is not a problem for normative ethics, it is a problem for psychological science.⁶

In psychological science, we make and test empirical claims about psychological phenomena. To test our hypotheses against reality, we need criteria for picking out the psychological phenomena we are interested in. Definitions provide such criteria. To see the distinction between definitions and empirical claims, imagine an evolutionary biologist who said: “Primate hands evolved for climbing.”⁷ This is an empirical claim about the evolutionary function of hands and not a definition. We can tell it is an empirical claim because we can picture an empirical reality that would have falsified the claim. We can imagine that primate hands evolved for cracking nuts, or some other purpose, and only later became instruments for climbing. Similarly, we can ask questions about how morality—or a certain element of morality—affects coexistence or survival. These, too, are empirical questions that we can address once we have identified the phenomena that count as moral.

Definitions allow us to test our claims. To test whether hands evolved for climbing, the biologist would need a way of identifying hands across evolutionary time regardless of their function. The biologist might define hands as “five bendable fingers and a palm at the end of each arm.” Then, the biologist could see whether primates developed hands because of the survival benefits of climbing or because of some other benefit, such as tool use. Analogously, if moral psychologists are to examine the social or evolutionary consequences of morality, we need a definition of morality that identifies psychological phenomena irrespective of their empirical consequences. We need a *psychological* technical definition of morality.

⁶The argument against defining morality by its consequences in psychological research is not an argument against consequentialism as a moral philosophy, nor is it an argument against other kinds of functionalist theories. The ethical tradition of consequentialism proposes that the morally right action is the action that leads to the best expected or actual consequences (Greene, 2014; Mill, 1879; Wiggins, 2006; Williams, 1973). Consequentialists readily grant the point that I have made in this section: It can be difficult to know the consequences of an act. But this is not a problem for consequentialism as a moral philosophy, because ethical consequentialism does not need to pick out a psychologically distinct set of phenomena that necessarily count as moral. The fact that we cannot always know what will generate the best consequences helps explain, for a consequentialist, why we have moral dilemmas. By contrast, a definition of morality *for psychological research* does need to pick out such a distinct set of psychological phenomena in order to study them. In psychological research, it would be unworkable if what counted as a moral phenomenon could only be determined long after it occurred.

⁷I thank a reviewer for suggesting this example.

(c) The Evaluating: Capturing the Set of (Morally) Right Actions

A third approach is to define morality in terms of right or good actions and traits. Much research on character development and moral identity takes this *evaluating* approach (Aquino & Reed, 2002; Krettenauer & Hertz, 2015; Nucci et al., 2014; Walker, 2014). These researchers define morality as the traits we ought to have—or that a plurality of research participants think we ought to have. In their influential study on moral identity, Aquino and Reed (2002) asked a sample of undergraduate students to list traits that “a moral person possesses” (p. 1426). Through content analysis, Aquino and Reed created a list of nine traits mentioned by at least 30% of their participants. (That the cut-off had to be 30% and not 100% hints at how differently people use the word “moral,” even people from same language community.) Their list, which became the basis for their Self-Importance of Moral Identity Questionnaire (SMI-Q), included the following nine traits: *caring, compassionate, fair, friendly, generous, helpful, hardworking, honest, and kind*.

After the publication of Aquino and Reed’s (2002) paper, numerous studies on moral identity have used the SMI-Q (for reviews, see Hertz & Krettenauer, 2016; Jennings et al., 2015). With time, researchers began to treat the nine traits of the SMI-Q as defining traits of morality. In their meta-analysis paper, Hertz and Krettenauer (2016) wrote that the SMI-Q “contains a list of nine attributes that are characteristic of a highly moral person” (p. 131). Here, the nine traits in the SMI-Q are no longer treated as traits that some of Aquino and Reed’s (2002) participants associated with morality. The traits become part of the scientific definition of a moral person, and individuals who possess these characteristics are said to have a “strong moral identity” (Hertz & Krettenauer, 2016, p. 129).

In a similar vein, research on moral disengagement tends to define morality based on what researchers evaluate as right and wrong behaviors (Bandura, 2002, 2016; McAlister et al., 2006). For instance, Bandura (1990, 2018) defines moral disengagement “a phenomenon in which moral self-sanctions are disengaged from detrimental behavior” (Bandura, 2018, p. 247). That is, to disengage morally means to engage in a “detrimental behavior” without viewing oneself negatively. Crucially, the literature on moral disengagement offers no separate criterion for determining which behaviors are detrimental. Rather, Bandura (2016) asserted that certain behaviors—capital punishment or efforts to increase birthrates in Western countries (pp. 392–393)—were inherently and self-evidently detrimental. He held these to be behaviors that a morally engaged person would condemn. By that logic, anyone who engages in such behaviors without feeling bad, and anyone who judges the actions to be okay, must have disengaged or “turned off” their morality (for discussion, see Dahl & Waltzer, 2018).

Kohlberg (1971) famously challenged those who defined morality as a list of desirable traits, colorfully calling such a list a “bag of virtues.” He wrote: “the trouble with the bag of virtues approach is that everyone has his own bag. The problem is not only that a virtue like honesty may not be high in everyone’s bag, but that my definition of honesty may not be yours” (p. 184). Indeed, reviews of lists of virtues across time, communities, and scholars reveal little consensus on which virtues are desirable (Nucci, 2004, 2019). In the face of disagreement, should morality be defined by the virtues that the researcher happens to prefer? Or the virtues preferred by some proportion of survey participants?

More fundamentally, to define morality via morally desirable traits traverses the old barrier between *ought* and *is* (Hume, 1738; Moore, 1903). Psychology, like any empirical science, is equipped to find out what *is* the case—what people actually do—but not what people *ought* to do. If psychological science defined morality as traits that are good or desirable, it would lack the means to determine which traits or actions actually fit within this definition.⁸ Even if most people in a community happened to believe that some trait constituted a virtue, it would not follow that this trait was actually (morally) good. The briefest reflection on societies that accepted slavery reminds us why we do not settle all moral questions by majority vote (Davis, 2006).

Although an empirical science cannot tell which actions are right and which are wrong, it can study how people *judge* which actions are right and which are wrong. Research on moral psychology investigates how some people come to support abortion rights while others oppose such rights (Craig et al., 2002; Dworkin, 1993; Gallup Inc., 2018; Helwig, 2006). This kind of investigation is the focus of the fourth approach to defining morality.

(d) The Normative: Capturing all considerations about right and wrong?

The *normative* approach is to define morality as comprising all normative considerations that generate judgments about right, wrong, good, or bad. We saw earlier how the Moral Foundations Questionnaire asks participants what they consider when they “decide whether something is right or wrong” (Graham et al., 2011; Haidt & Graham, 2007). In a similar spirit, Haidt (2001) and Schein and Gray (2018b) defined moral judgments as any evaluations based on virtues taken to be obligatory by a culture or subculture. And Hamlin (2013) defines our moral sense as a “tendency to see certain actions and individuals as right, good, and deserving of reward, and others as wrong, bad, and deserving of punishment” (p. 186; see also Malle, 2021). Here, too, morality is defined to incorporate all judgments of right, wrong, right, or bad, regardless of subject matter.

There is nothing wrong with studying norms—many people do (Bicchieri, 2005; Heyes, in press; Legros & Cislighi, 2020; Schmidt & Tomasello, 2012; Sripada & Stich, 2006). But those who do “norm psychology” usually treat morality either as a subcategory of norms or a category altogether different from norms. Laying out their framework for the psychology of norms, Sripada and Stich (2006) write that the “category of moral norms is not coextensive with the class of norms that can end up in the norm database posited by our theory” (p. 291). To exemplify non-moral norms, they mention “rules governing what food can be eaten, how to dispose of the dead, how to show deference to high-ranking people” (p. 291). In her influential treatise on social norms, Bicchieri (2005) writes that “[s]ocial norms should also be distinguished from moral rules” (p. 8). While these authors do not agree on where to draw the line, they agree that there should be a line between morality and (other) norms.

⁸My argument does not imply moral relativism. As human beings, we all judge that some acts are morally right and other acts are morally wrong—whatever we mean by “moral.” But as psychologists, engaged in empirical science, we have no special tools for deciding which human actions are right and which are wrong. Psychology cannot prove that we ought to donate 10% of our income to charity, just as economics cannot prove that we ought to arrange our societies so as to maximize individual freedom (Friedman, 1962), and just as biology cannot prove that we ought to genetically modify human embryos.

For if we define morality to include all normative considerations, we blur some common-sense distinctions. People make normative judgments about right and wrong about almost any human activity (Daston, 2022). As Kohlberg (1971) observed, there is a right and a wrong way of mixing a martini, training for a sports race, and playing solitaire. And people do judge violations of instrumental norms or board game rules differently from how they judge harming or stealing from others (Dahl & Schmidt, 2018; Dahl & Waltzer, 2020a; Turiel, 1983). For instance, most adults think it is okay to violate norms of instrumental rationality (e.g., how to train for a race, how to bake a cake) if the person does not care about the goal served by that norm (e.g., doing well in the race, Dahl & Schmidt, 2018). In contrast, they judge that it is generally wrong to harm another person, whether or not you care about harming them (Dahl & Schmidt, 2018; Knobe, 2003). It is up to you whether you should try to do well in the marathon you signed up for. It is not up to you whether you should respect the welfare and rights of other people.

One feature that sets most game rules apart from prototypical norms is alterability. Children and adults view the rules of board games or sports as somewhat arbitrary and, accordingly, somewhat alterable (Dahl & Waltzer, 2020a; Hardecker et al., 2017; Köymen et al., 2014). They imagine that alternative rules would do equally well. Wikipedia currently lists 26 invented and regional variants of checkers (“Checkers,” 2022). In one variant—“Russian Column Draughts”—captured pieces are not removed but placed under the capturing piece to form a tower. Even if Russian checker players preferred Russian Column Draughts, they would probably recognize that alternative variants can be just as fun. In contrast, many standard moral norms, like the prohibition against hitting or stealing, are not seen as alterable. Even if the government legalized stealing or unprovoked violence, most people would judge such actions wrong (see Killen & Smetana, 2015; Turiel, 2015).

I will return to the distinction between moral and non-moral judgments in the next section. My point here is that, in ordinary language, many normative judgments about right, wrong, good, and bad fall outside what most people call “morality.” Some divergence between scientific and everyday usage is unavoidable—we already abandoned the idea of complete overlap. But do pool moral judgments with all normative judgments is like pooling the whales with the fishes. We seek a technical definition that overlaps with everyday usage yet retains enough psychological unity to theorize over. For this reason, we will seek a technical definition that sets moral judgments apart from at least some other normative judgments. This way, our technical definition will not force us to say that judgments about mixing martinis or playing checkers are moral judgments, and our theories of morality will not have to account for those judgments. We seek a definition that renders moral considerations *distinctive* among normative considerations

Summary: Four Approaches to Defining Morality and Their Difficulties—I have discussed four approaches to defining morality (Table 1). The linguistic approach defines morality as what people call “morality.” The functionalist defines morality in terms of its external function or consequences. The evaluating defines morality as a set of right or good characteristics. And the normative approach defines morality to incorporate all considerations about right, wrong, good, or bad. My goal was to provide a taxonomy of

common and intuitive approaches to defining morality—of things we could do when we define morality for psychological research.

On inspection, each approach revealed its rubs. The linguistic approach faced the difficulty that the word “morality” has sundry uses, within and between individuals. To cope, we need a technical definition. For the functionalist approach, the trouble was we could not know whether something counted as moral without knowing its consequences, which could be unknowable or not knowable for a long time. When studying a psychological phenomenon, we need a definition that refers to psychological characteristics. The evaluating approach had the downside that empirical psychological science lacks a method for identifying (morally) good or bad traits or acts. Empirical science needs a descriptive definition of morality. And the normative approach failed to distinguish morality from other normative considerations, which would deviate unnecessarily from ordinary usage—and from current scientific usage. We will look for a definition that distinguishes morality from at least some normative considerations. In the next section, I will propose a definition that meets these four criteria.

What to Do When We Define Morality: A Proposal

When we define morality for psychological science, we can look for a definition that is (a) *technical* in that it overlaps with some, but not all, everyday uses of the words “moral” and “morality”; (b) *psychological* in that it identifies widely shared psychological phenomena without reference to their external functions or consequences; (c) *descriptive* in that it captures people’s evaluative judgments without committing scientists themselves to making evaluative judgments; and (d) *distinctive* in that it separates moral considerations from at least some other kinds of considerations about right and wrong.

Guided by these four criteria, I propose to define morality as obligatory concerns with how to treat other sentient beings (i.e., with others’ welfare, rights, fairness, and justice), as well as the reasoning, judgments, emotions, and actions that spring from these concerns (Dahl, 2019; Killen & Smetana, 2015; Turiel, 1983, 2015; Turiel et al., 2018). For instance, a judgment about hitting is moral if it springs from an obligatory concern with others’ rights or welfare; the judgment is non-moral if it springs from a concern with staying out of trouble. An angry outburst at injustice is moral if it reflects the emoter’s obligatory concern with fairness or justice; the anger is not moral if the person is merely angry about not getting what they wanted (Batson et al., 2009; Wakslak et al., 2007).

The definition is not the sole sensible definition of morality for psychological research. Others may adopt different definitions that serve other ends. In this respect, my attitude is like Wittgenstein’s: “Say what you choose, so long as it does not prevent you from seeing the facts. (And when you see them, there is a good deal that you will not say)” (Wittgenstein, 1953, p. 37). If another definition suits a researcher’s project better—and allows the researcher to better “see the facts”—researchers could state that definition and their reasons for adopting it. Or if psychologists come together to develop a different, shared definition of morality—as Grossman and colleagues (2020) did for wisdom or as Scherer and Mulligan (2012) did for emotion—that would be a welcome development.

In the meantime, each researcher can advance the field by explaining which definition of morality they adopt and why they adopt it. What follows is my attempt.

Morality as Rooted in Obligatory Concerns With Others' Welfare, Rights, Fairness, and Justice

My definition of morality, rooted in obligatory concerns with how to treat others, builds on the work of Turiel (1983, 2015) and his collaborators (Killen & Smetana, 2015; Nucci, 2001).⁹ I single out these four types of obligatory concerns because they are inherently about how to treat *other sentient beings*, be they individuals, social groups, or animals. Concerns with *others' welfare* are about promoting and protecting the psychological and physical well-being of others (Dahl et al., 2018; Schein & Gray, 2018b). Concerns with *rights* are about the protections and entitlements people owe to each other (Helwig, 2006; Hunt, 2008). Concerns with *fairness* are about distributing resources and privileges among others (Blake et al., 2015; Killen et al., 2018). And concerns with *justice*, as I define them here, are about the rewards and punishment that others deserve (D. Miller, 2017; Turiel et al., 2016).

Philosophers and psychologists have proposed both formal and substantive properties that distinguish morality from other normative domains (Gert & Gert, 2017; Kohlberg, 1971; Stich, 2018; Turiel, 1983). Formal properties are properties that can apply regardless of content. One common formal criterion is universalizability: A norm is moral if you can want it to be a universal law that every one followed (Kant, 1785; Kohlberg, 1971; Levine et al., 2020). This criterion imposes no limits on what those norms are about. We could imagine universalizing any norm, even if there are many norms we would not want to universalize. A person could treat a norm that everyone play Russian Column Draughts as a moral norm—as long as they could will that norm to become a law for law for all of humanity.

Substantive properties are properties that refer to specific content. A requirement that moral judgments always be about the relation between an agent and a suffering patient is a substantive requirement (Gray et al., 2012). This substantive requirement would preclude norms about how to play solitaire from being counted as a moral norm.

My definition includes both formal properties (the obligatoriness of concerns) and substantive properties (the concerns with others' welfare, rights, fairness, and justice). I will first clarify the notion of obligatory concerns. Next, I will discuss why I separate (moral) obligatory concerns with others' welfare, rights, fairness, and justice from other (non-moral) obligatory concerns.

Obligatory Concerns—The notion of concern, common in theories of emotion, is a cognitive-motivational construct (Dahl et al., 2011; Frijda, 1986; Lazarus, 1991; Mulligan & Scherer, 2012). Concerns are cognitive representations that we care about, such as human welfare, the will of God, or the success of our sports team (Dahl, 2017b; M. Lewis, 2007).

⁹In his influential 1983 book, Turiel defined the moral domain in terms of “prescriptive judgments of justice, rights, and welfare pertaining to how people ought to relate to each other” (p. 3, also pp. 35–44). Expanding Turiel’s definition, the present definition explicitly incorporates concerns, emotions, and actions to incorporate research on early moral development and the role of emotions from the past 20 years.

Individuals map concerns onto specific situations or issues (C. S. Carver & Scheier, 2001). When we see one person hit another, we perceive a connection between the event and our concern for the welfare of others. When we watch our team play, we map our concern for the team's success onto goals for our team (good news) and goals against our team (bad news).

Obligatory concerns are concerns that it would be wrong *not* to have (Dahl et al., 2020; Dahl & Schmidt, 2018). Obligatory concerns are those we believe others could not “reasonably reject,” if we borrow Scanlon's (1998, p. 189) phrase.¹⁰ If you hold that everyone ought to concern themselves with the welfare of others, and that it would be wrong to lack this concern, then you would be treating that concern as *obligatory*. We can distinguish obligatory concerns from non-obligatory concerns, such as personal preferences, by two features: obligatory concerns, unlike non-obligatory concerns, can give rise to judgments that are both categorically negative and agent-neutral (Dahl & Freda, 2017; Kohlberg, 1971; Scanlon, 1998; Tomasello, 2018). Let me pause briefly to unpack each notion.

To consider a concern obligatory, a person must judge that it is *categorically wrong* to lack this concern, not merely suboptimal or non-preferred. Someone might wish others shared their love for red wine or soccer (Nucci, 1981, 2001). Still, unless they viewed concerns with red wine and soccer as obligatory, they would not view the indifference to red wine or soccer negatively. They would not experience frustration or disappointment upon learning that others preferred white wine or basketball. In contrast, if others were wholly unconcerned with women's right to abortion, and the person viewed the concern with a right to abortion as obligatory, the person would condemn this lack of concern.

To demonstrate the categorical judgments that evidence obligatory concerns, we need categorical assessments. We cannot, for instance, use measurements of preferences to tell which concerns a person views as obligatory and which they view as non-obligatory. In research on the infant precursors of morality, one common paradigm is to show infants puppet shows involving three puppets: a helpful puppet, a hindering puppet, and a recipient puppet (Hamlin, 2013; Hamlin et al., 2007; Margoni & Surian, 2018). The helpful puppet always helps the recipient puppet, for instance, by retrieving a ball that the latter dropped; the hindering puppet always hinders the recipient puppet, taking the dropped ball instead of returning it. In a subsequent assessment, infants have the choice of reaching toward the helpful or hindering puppet. Most studies find that most infants reach toward the helpful puppet, suggesting that infants prefer helpful over hindering puppets. But such a preference for helpful over hindering puppets does not demonstrate categorically negative evaluations of the sort that obligatory concerns give rise to. The preference for the helpful puppet does show that infants had a categorically negative view of the hindering puppet; you can prefer Rome over Paris and still adore Paris (Dahl & Waltzer, 2020b).

¹⁰It would go beyond the scope of this article to further detail what makes concerns seem obligatory. For relevant discussions, the reader may consult deontological as well as consequentialist perspectives on moral philosophy (e.g., Kant, 1785; Korsgaard, 1996; Mill, 1879; Rawls, 1971; Scanlon, 1998; Wiggins, 2006). For the present purposes, it will suffice to say that individuals *do* deem some concerns to be obligatory. Saying that a concern is obligatory is different from saying that an action is obligatory. Even if people think we are always obligated to concern ourselves with the welfare of others, they do not judge that we are always obligated to help others or to otherwise prioritize others' welfare in all our actions.

The second characteristic of obligatory concerns is agent neutrality. Agent neutrality means that a person will condemn the lack of obligatory concerns regardless of their own role in the situation (Nagel, 1989; Scanlon, 1998; Tomasello, 2018). When you deem a concern obligatory, you will demand it from a first-, second-, and third-person point of view. Let us stipulate that a person who engages in unprovoked violence is unconcerned with the welfare of others. Let us also stipulate that I deem the concern with others' welfare as obligatory. Then, I would reprehend unprovoked violence regardless of whether I am hitting someone else (first person), whether someone else is hitting me (second person), or whether I observe one person hit another (third person). If I merely thought it was bad for others to hit me, but completely fine for me to hit others, I would not be viewing the concern with others' welfare as obligatory.

Third-person evaluations set humans apart from non-human primates. Chimpanzees do not like to be hit by others (second person), and they often refrain from violence against kin (first person). Still, chimpanzees do not generally react negatively when they observe unrelated individuals hitting each other (Silk & House, 2011; Tomasello, 2016).

Agent neutrality does allow moral judgments to consider social roles (Korsgaard, 1996). We can make agent-neutral evaluations about doctors' duties toward their patients and about parents' duties toward their children. What agent neutrality precludes is that the person making the evaluation takes into account which role they—the evaluator—have in the situation. Imagine that I say that a soldier is obligated to concern themselves with the commands of an officer. I am saying that the soldier has to concern themselves with the officer commands irrespective of whether I am the soldier, the officer, or a bystander. I am not thereby saying that everyone, even civilians, has to concern themselves with the officer commands. Obligatory concerns can be specific to a social role or relationship; but they are not specific to me as me, nor to you as you.

Obligatory concerns do give rise to judgments other than judgments of obligation. Obligatory concerns can give rise to judgments about which of two permissible actions is *better*. And obligatory concerns generate judgments that an act is supererogatory: good, but not required (McNamara, 2011). People often judge some acts of helping others as supererogatory, especially when the helping would be costly to the helper (Dahl et al., 2020; Kahn, 1992; Killen & Turiel, 1998; J. G. Miller et al., 1990). Many of these supererogatory judgments about helping spring from obligatory concerns with others' welfare: Helping would be a good thing to do because it benefits the welfare of others (Dahl et al., 2020).

Judgments of supererogation illustrate a broader characteristic of obligatory concern. To treat a concern as obligatory is to hold that the concern should always be *considered*, not that it always should be *prioritized* (Dahl & Schmidt, 2018; Scanlon, 1998). In dilemmas that pit two obligatory concerns against each other, it would actually be impossible to prioritize both concerns (Dahl et al., 2018; Turiel, 2002; Turiel & Dahl, 2019). If the rights of one person conflicts with the welfare of others, you cannot prioritize both your concerns with the former's right and your concerns with the latter's welfare. You can still think it obligatory to consider both concerns as you make your judgment.

And even after the judgment, the down-prioritized concerns continue to influence how we think and feel about the situation. When people harm others, they can still experience conflict and guilt about their decisions, even if they deem the harm justified (Hoffman, 2000; Wainryb, 2011). People who judge that they should disclose fraudulent practices in their company, and decide to blow the whistle, are often torn. Their actions prioritize their obligatory concern with honesty, but it violates their obligatory concerns with loyalty to their company and their co-workers. Whistleblowers can therefore feel deeply conflicted about their actions (Alford, 2002; Waytz et al., 2013). Other examples of dilemmas and conflicts are more mundane. Since each person's obligatory concerns form an incoherent whole, we often encounter situations that pit one concern against another. To make sense of those conflicts, we will want to draw some distinctions among kinds of obligatory concern, which I will consider next.

Separating Obligatory Concerns with Sentient Beings From Other Obligatory Concerns

Concerns—Moral concerns are a subset of obligatory concerns, namely obligatory concern with the promotion and protection of others' welfare, rights, fairness, and justice. Unlike other obligatory concerns, these deal intrinsically with how to treat other sentient beings. Whenever we consider others' welfare, rights, fairness, and justice, it is *somebody's* welfare, rights, fairness, and justice, whether that somebody is an adult, a child, an animal, or a social group.

People need not have a specific victim in mind when they map a situation onto moral concerns.¹¹ To meet my definition of a moral judgment, a judgment needs only to deem that a transgressor showed insufficient concern for others' welfare, rights, fairness, or justice, whether or not the transgressor harmed a specific victim. Within my framework, a person may judge that racist slurs are (morally) wrong even if nobody hears the slurs; and a person may judge a deceased person has (moral) rights to respectful treatment without assuming that the deceased person is harmed in an afterlife (Jacobson, 2012).

Thus defined, moral concerns differ from other obligatory concerns, such as concerns with authority commands. As we saw earlier, most people deem that we ought to concern ourselves with the demands of legitimate authorities, even if people disagree about which authorities are legitimate (Frimer et al., 2014; Laupa, 1991). Religious individuals also view concerns with gods or other religious authorities as obligatory entities (Nucci & Turiel, 1993; Srinivasan et al., 2019). In contrast to moral concerns, concerns with authority commands regulate social as well as non-social actions. Some authority concerns do tell us how to act toward another person (e.g., military officers requiring soldiers to salute them). But unlike moral concerns, authority concerns can also tell us how to promote our *own* wellbeing, as when health authorities tell us to smoke less and exercise more.

As Figure 2 shows, I classify concerns with authorities as a type of *conventional* concern (Dahl & Waltzer, 2020a; Killen & Smetana, 2015; Turiel, 1983, 2015), along with concerns

¹¹For my definition, there is nothing puzzling about the so-called "harmless moral wrongs" that have caused such debate among moral psychologists (Gray et al., 2014; Haidt et al., 2000; Royzman et al., 2015): The perception of actual harm to a specific individual is only one of the things that could lead someone to map an event onto a moral concern, as I define them here. I thank a reviewer for requesting this clarification.

with traditions and consensus. Conventional concerns are obligatory concerns with the requirements of social institutions, be it the authorities atop those institutions (e.g., requests from a president or boss), the traditions that characterize the institutions (e.g., customs for how to dress in the office or how to behave a ceremony), or the consensus that the institution produces (e.g., the laws passed by a parliament or the agreement among checker players on which rules to use before a game).

A common misconception is that the label “conventional” implies lower importance compared to “moral”. Critics of the moral-conventional convention have asked why some norms should be relegated to “mere conventions” (Haidt & Graham, 2007, p. 100)? There is nothing “mere” about conventions, as I talk about them here. I use “moral” and “conventional” in a descriptive sense that implies no ranking of their importance. Individuals sometimes judge that they should violate a moral concern to prioritize a conventional concern, as when soldiers judge that they should follow an officer’s orders to harm another (Osiel, 2001; Turiel, 2002). From an empirical point of view, there is nothing inherently irrational or wrong about prioritizing conventional concerns over moral ones.

The technical distinction between moral and conventional concerns is useful because, like whales and fish, they tend to operate differently (see Killen & Smetana, 2015; Turiel, 1983, 2015). For instance, children and adults treat norms that serve (moral) concerns with others’ welfare, rights, justice, and fairness as more generalizable and less authority-dependent than norms that serve conventional concerns. Most judge that unprovoked violence would be wrong even in a place that had no rule against unprovoked violence (generalizability) and even if authorities gave permission (authority independence). In contrast, most would judge that it is okay to wear casual clothes to an office that has no dress code and that wearing casual clothes to the office is okay if the company C.E.O. gave permission.

People treat moral concerns as more generalizable and less authority-dependent than conventional concerns because moral concerns map onto intrinsic features of interpersonal relations (Turiel, 1983). Hitting another tends to cause pain to a victim no matter what the context is and no matter what authorities—even gods—permit (Nucci & Turiel, 1993; Srinivasan et al., 2019). Taking something that already belongs to another violates that person’s property rights, no matter where you are and no matter what authorities permit. (Of course, there is tremendous contextual variability in the procedures for *establishing* ownership rights; but the violation presupposes that ownership is established.) In contrast, showing up to the office in sweatpants has no intrinsic connection to others’ welfare: How your clothing affects others depends on how your clothing relates to prevailing dress codes and other customs.

To prevent misinterpretation: I do not treat the patterns of generalizability and authority independence as *defining* features of morality (Kelly et al., 2007; Machery & Stich, 2022; Stich, 2018). The patterns are correlates of morality. Even concerns about others’ welfare, rights, justice, and fairness are somewhat sensitive to authorities and other contextual factors (Corzine & Dahl, 2022; Kelly et al., 2007; Rhodes & Chalik, 2013). In contact sports, views about which tackles are celebrated and which are wrong depend partly on what the rules of the sport allow (Bredemeier & Shields, 1986). And even non-moral rules can sometimes

be generalized and authority-independent. Some religious fundamentalists believe that the commands of their god should be followed by everyone; in other words, they generalize their religious commands to apply to all contexts (Altemeyer & Hunsberger, 2004; Emerson & Hartman, 2006). But this does not mean that religious individuals fail to distinguish concerns with welfare, rights, fairness, and justice from concerns with religious authorities. Most religious youth judge that unprovoked force would be wrong even if their god had never prohibited it (Nucci & Turiel, 1993; Srinivasan et al., 2019).

Rather than being defining, the generalizability and authority independence associated with moral concerns furnish a key rationale for distinguishing between moral and other obligatory concerns. The distinction between moral and non-moral concerns can help us explain variance in people's views about which rules should be generalized and which can be altered. By analogy, the SARS-CoV-2 virus often causes COVID-19 symptoms and transmission to others. But symptoms and transmission are not defining features of the virus. The SARS-CoV-2 has a genetic definition that allows detection even in the absence of symptoms or transmission. But the symptoms and transmission of illness furnish a rationale for distinguishing SARS-CoV-2 from other viruses. If SARS-CoV-2 tended to operate exactly like SARS-CoV-1, we would not care much about the distinction between the two viruses, even if their genetic makeup differed slightly. For morality, as for viruses, the definition is deterministic but the rationale for creating that definition is probabilistic.

Moral concerns also differ from *prudential concerns* with the agent's own welfare (Nucci et al., 1991; Tisak & Turiel, 1984). Prudential concerns can lead us to judge that another person should eat more healthily, refrain from dangerous activities, or leave an unhealthy relationship. Like moral concerns, prudential concerns are about intrinsic features of situations. Mountain climbing alone without ropes and bolts—known as climbing *free solo*—is dangerous no matter which country you are in and no matter what authorities prohibit or permit.

Still, the application of prudential concerns tends to differ from the application of moral concerns. For instance, people usually grant individuals more liberty to risk their own welfare than to risk the welfare of others (Dahl & Schmidt, 2018). Children judge it worse to push another person off a moving swing than to jump off the swing yourself (Tisak, 1993). Conversely, people also judge that individuals have the right to prioritize their own welfare over others—sometimes called “agent-favoring prerogatives” (McNamara, 2011). If we are both hungry, and I brought lunch when you did not, most would judge that I had the right to eat my lunch, even if sharing would be kind (Dahl et al., 2020; J. G. Miller et al., 1990). In short, people grant each other more leeway both to down-prioritize prudential concerns (by climbing free solo) and to (up-)prioritize prudential concerns (by eating my lunch) relative to comparable moral concerns.

The distinction between moral and non-moral obligatory concerns has led some to infer a corresponding distinction between moral and non-moral rules or situations (for discussion, see Dahl & Waltzer, 2020a). However, the distinction between morality, conventionality, and prudence is neater at the level of concerns than at the level of situations. A given event can be mapped onto both moral and non-moral concerns, and two people can map

situations onto concerns differently (Turiel, 1983, 1989; Turiel et al., 1991). When an officer commands a soldier to harm an enemy, the soldier may map the act onto conventional concerns with authority commands, moral concerns with the enemy's welfare and rights, or both (A. P. Fiske & Rai, 2014; Wainryb, 2011). These mixed events do not invalidate the distinction between moral and non-moral concerns. On the contrary, such mixed events evidence the usefulness of distinguishing between moral and non-moral concerns (Killen & Dahl, 2021; Smetana, 2013; Turiel, 1989). Only by considering the competing concerns that individuals strive to coordinate can we account for the conflict and uncertainty that the dilemmas evoke.

Two Sources of Individual and Cultural Variability—People vary in how they apply their moral and non-moral obligatory concerns to specific issues and situations. Even people who share moral and other obligatory concerns can form opposing judgments about a specific issue. Dworkin (1993) observed that, underneath their disagreements about abortion, pro-life and pro-choice advocates share concerns with the sanctity of life and women's rights to decide over their own bodies. Where the pro-lifers and pro-choicers differ is how they apply those concerns to the question of whether a mother can terminate a pregnancy.

Two major processes can explain why persons with similar concerns can form different moral judgments: factual beliefs and coordination of competing values (Dahl & Killen, 2018b; Turiel, 2002, 2015). This is not the article to offer a complete account of variability in views about right and wrong. Still, a brief review of these two sources of variability will show how my proposed definition of morality allows for developmental, individual, and cultural variability.

Factual beliefs. The mapping of concerns onto specific events relies on so-called “informational assumptions” or factual beliefs (Asch, 1952; Turiel et al., 1987, 1991; Wainryb, 1991). Factual beliefs are descriptive (i.e., not evaluative) beliefs about how the physical, social, or supernatural world works. For instance, parents have factual beliefs about whether corporal punishment helps children learn or causes lasting harm to children (Gershoff, 2002; Wainryb, 1991). These beliefs lead parents to judgments about whether corporal punishment is okay. Parents who believe that corporal punishment helps children learn make more positive evaluations of corporal punishment than those who see corporal punishment as mostly harmful. Henrich and Henrich (2010) found that many mothers on Fiji believed that certain seafoods were harmful to a fetus. This informational assumption mapped the act of eating seafood while pregnant onto moral concerns with the welfare of a fetus (Dahl & Waltzer, 2020b; Turiel et al., 1987). Perceiving this mapping, many of the Fiji participants judged it wrong for pregnant or nursing women to eat those kinds of seafood. A person who did not perceive those mappings of seafoods onto fetal harm would, expectably, find it perfectly acceptable for a pregnant woman to eat those seafoods. In this way, different factual beliefs can lead to different judgments, even among people with similar moral concerns.

Coordination of Competing Concerns. By coordination, I mean efforts to manage multiple concerns that map onto the same event (Campos et al., 2011). Coordination is common because many—perhaps most—events can map onto multiple concerns (Turiel,

1983, 1989; Turiel & Dahl, 2019). Abortion maps onto both the moral standing of the fetus and the mother's right to choose. Blowing the whistle on illicit practices in your company maps onto concerns with loyalty, honesty, and the well-being of the whistleblower. Speeding maps onto concerns with traffic laws, others' rights and welfare, and one's own welfare.

When people map competing concerns onto an event, they strive to align, balance, or prioritize those concerns (Dahl & Killen, 2018b; Turiel et al., 1991; Turiel & Dahl, 2019). A soldier commanded to harm a civilian must choose between disobeying an officer and causing harm to another person (Bandura, 2016; Browning, 1992; Doris & Murphy, 2007). Most soldiers will generally hold both that they ought to obey authority commands and protect the welfare of innocent civilians, and they will therefore struggle when the two concerns conflict. As they strive to coordinate these competing concerns, people may experience deep conflict, consider reasons for both courses of action, and experience regret no matter which course they take (Turiel & Dahl, 2019; Wainryb, 2011). Faced with difficult dilemmas, people may opt to coordinate their competing concerns in different ways. Some soldiers reject orders to harm civilians, while others decide to accept the orders, even if they experience the same underlying conflicts between obligatory concerns (Browning, 1992).

Explaining and predicting differences in coordination is an exciting scientific endeavor (see e.g., Holyoak & Powell, 2016; Nucci et al., 2017). For the present purposes, the central point is this: Two people can map the same obligatory concerns onto a situation but differ in how they coordinate those competing concerns.

From Moral Concerns to Moral Reasoning, Judgments, Emotions, and Actions

—When we define morality using the cognitive-emotional concept of concerns, we avoid the troublesome separation of reason and emotion (Greene, 2008; Lazarus, 1991). According to standard theories of emotion, concerns give rise to emotions via cognitive *appraisals*—cognitive perceptions of how events relate to our concerns (Frijda, 1986; Lazarus, 1991; Moors & Scherer, 2013). Argentinian soccer fans watching the 2022 World Cup Final between Argentina and France rejoiced in the 20th minute of the game, when the referee awarded a penalty kick to Argentina. The French felt dread and sadness. Both groups of fans appraised the penalty to mean that Argentina would likely score. (In soccer, 85% of penalty kicks lead to goals.) Argentinian fans appraised an event (awarded penalty kick to Argentina) to increase the chance that their concern (Argentinian victory) would be satisfied—a source of joy. French fans appraised the same event to *decrease* the chance of their concern (French victory) being satisfied—a source of fear and sadness.

Appraisals work for moral concerns as they do for sportive concerns. When we are concerned with someone's welfare and appraise some event as posing an immediate threat to that person, we tend to have an emotional reaction. The nature of that emotional reaction depends on our appraisal. If we appraise the threat as a natural disaster, we may experience fear; if we appraise the threat as an intentional effort to harm the person, we may experience moral outrage (Hoffman, 2000; Hutcherson & Gross, 2011). Within this framework, any emotional reaction—moral or non-moral—involves an appraised relation between ongoing events and the “fate of our concerns,” as Frijda (1986, p. 334) put it.

We can call judgments, reasoning, emotions, and actions *moral* whenever they spring from moral concerns. The same motoric movements could be a moral or non-moral actions, depending on whether a moral concern guided the action (Nucci, 2004). An act of helping could spring from a non-moral concern, as when a person helps another steal money for financial benefit (e.g., a bank clerk helping a bank robber to get a cut). And an act of harming could spring from a moral concerns, when a person harms another to save a life (e.g., a surgeon, or someone who uses violence in self-defense. Dahl et al., 2020; Kohlberg, 1971; Nucci, 2004).

Moral concerns guide many of our judgments, reasoning, emotions, and actions. Children, adolescents, and adults judge that violence against others is wrong because it causes harm and violates rights (Dahl & Freda, 2017; Killen & Smetana, 2015; Nucci et al., 2017; Wainryb et al., 2005). These are moral judgments, involving moral reasoning, insofar as they derive from moral concerns, for instance obligatory concerns with others' welfare. Moral judgments can be constitutive of moral emotional reactions, such as moral outrage at violence, or guide moral actions, such as interventions to stop violence (Batson et al., 2009; Cushman et al., 2012; Hoffman, 2000). In each case, the classification of a phenomenon as moral derives from the involvement of underlying moral concerns with others' welfare, rights, fairness, and justice.

The explicit inclusion of emotions and actions in the definition of morality sets my definition apart from strictly cognitive definitions of morality (Kohlberg, 1971; Piaget, 1932; Turiel, 1983). Kohlberg (1971) wrote that “the basic referent of the term *moral* is a type of *judgment* or a type of *decision-making process*, not a type of behavior, emotion, or social institution” (p. 169). Although these prior authors were also interested in the role of emotion in morality, their definitions did not provide a ready answer to questions about which emotions or actions counted as moral. Considering the growth of affective science over the past 40 years, and the continued interest in relations between moral judgments and actions, I see the inclusion of emotion and action in a definition of morality as a strength (Barrett et al., 2016).

Is My Definition WEIRD and Liberal?—Definitions of morality in terms of others' welfare, rights, fairness, and justice have been criticized for liberal and Western bias (Haidt, 2013; Machery, 2018). Similar critiques have been leveled at other areas of psychology that, according to Henrich and colleagues (2010), have focused on samples and perspectives that are WEIRD—Western, Educated, Industrialized, Rich, and Democratic. When directed at the present topic, the question becomes whether my definition reflects a cultural bias that overlooks the perspectives of underrepresented communities.

Haidt (2013) has argued that a definition of morality in terms of others' welfare, rights, fairness, and justice represents is biased even within WEIRD countries. He has argued only political liberals restrict morality to issues of others' welfare, rights, fairness, and justice. In the words of Haidt and Graham (2007), “conservatives have moral intuitions that liberals may not recognize” (p. 98). They argued that conservatives and non-WEIRD individuals, unlike WEIRD liberals, see morality as also including concerns with authorities, loyalty, and purity. The root of this liberal bias, Haidt (2013) suggests, is the political orientation of the

researchers themselves: “Nearly all moral psychologists are politically liberal—I know of none who self-identify as conservative. So the moral worldview of US conservatives can seem at times quite alien and undeserving of respect” (p. 291).

In my discussion of cultural similarities in values, I noted that all communities treat concerns with authorities, loyalty, and purity as obligatory. Communities can differ in which authorities they deem legitimate and which groups they see as requiring loyalty. But there are no human groups—liberal or conservative, WEIRD or non-WEIRD—who think we should respect no authorities, be loyal to no groups, and never care about what is disgusting (Frimer et al., 2014; Graham et al., 2013; Turiel, 2015). I have already discussed one example of such evidence from Haidt and Graham (2007), who reported that even U.S. liberals rated concerns with authorities, loyalty, and purity as being between “slightly relevant” and “somewhat relevant” to their judgments about right and wrong *in general*. And liberals would surely rate the commands of certain authorities “highly relevant” for their judgments about certain issues, like the prescriptions of doctors about which medications to take (Frimer et al., 2014). So—as I noted in my discussion of the *normative* approach to definitions—to define concerns with authorities, loyalty, and purity out of the moral domain does not imply that these concerns are irrelevant to judgments about right and wrong.

Machery (2018) suggests even more radical critique of my definitions like mine. Machery questions the whole idea of distinguishing morality from other normative considerations. He argues that morality is a modern Western invention. Machery reviews linguistic evidence that many non-English languages lack equivalent terms for “right,” “wrong,” and “moral.” He also cited unpublished evidence that, when asked whether a judgment they had made was a “moral” judgment, Indian participants did not separate “moral” from “non-moral” judgments the same way U.S. participants did (for related findings, see Levine et al., 2021). If morality is indeed a historical invention, Machery (2018) suggests, “Westerners’ moral emotions and their moral motivation may not be justified” (p. 264).

In my discussion of the linguistic approach to defining morality, I wrote that the variability in what people call “moral” does not render a technical definition of morality useless. Even variability in whether languages have a term like “moral” does not render a technical definition useless. Finding that some languages count whales as fishes, or have no word for “cell,” would not invalidate the scientific definitions of “fish” and “cell.” On the contrary, linguistic variability renders a technical definition necessary for scientific study. Only a technical definition lets us determine and communicate what we seek to explain, and what we do not seek to explain, within and beyond the scientific community.

Still, a critic might wonder if researchers are only interested in obligatory concerns with others’ welfare, rights, fairness, or justice because they are liberals from WEIRD countries. I do not know exactly why I became interested in morality thus defined. I do know that people in virtually all communities share those obligatory concerns with how to treat other sentient beings, which operate alongside obligatory concerns with authorities and concerns that fall outside my technical definition of morality. My identity and background—whatever those happen to be—cannot explain my interest in what I define as morality since I share that interest with people of all identities and backgrounds. (It reiterate that my calling a

concern “moral” in the scientific sense has no implications for how that concern ought to be prioritized. I continue to use “moral” in a strictly descriptive sense.) There is simply nothing (uppercase) WEIRD about treating concerns with others’ welfare, rights, fairness, and justice as important and obligatory. And there would be something (lowercase) weird about being indifferent to concerns with authorities, loyalty, and purity—weird since no communities exhibit such indifference. In this light, my definition of morality looks neither WEIRD nor weird.

Looking Back: How the Proposed Definition Meets the Four Criteria—I have now proposed a definition of moral concerns, moral judgments, moral emotions, and moral actions. This definition can form the basis of a testable theory of morality. The definition is not itself such a theory. It merely serves to identify the phenomena that a theory of moral psychology would seek to explain. Emphasizing the distinction between a definition and a testable theory, I have made this paper about the former and not the latter.

Having elaborated on my proposed definition, I now consider whether the definition meets the criteria set out earlier in the article and summarized in Table 1. It was these criteria that led us to abandon the four other approaches to defining morality (the *linguistic*, *functionalist*, *evaluating*, and *normative*) and seek another approach—one I characterized as *technical*, *psychological*, *empirical*, and *distinctive*.

(a) Technical: The definition should overlap with some, but not all, common uses of the word “moral.”: Our technical definition of morality, in terms of obligatory concerns with others’ welfare, rights, fairness, and justice, places itself squarely within ordinary and scholarly usage of the word (Ellemers et al., 2019; Gert & Gert, 2017; Graham et al., 2011; Turiel, 1983; Wiggins, 2006). Condemnations of violence, endorsements of human rights, and fights for social equity are all prototypical moral issues that fall under my proposed definition of morality. I cannot readily think of an example of something that falls under the proposed definition of morality that is *not* labelled “moral” in common usage (Figure 1; although a few counterexamples might be found). Of course, the converse does not hold: There are many phenomena that some might label “moral” that fall outside the definition I propose, such as respect for religious or secular authorities (Graham et al., 2013; Levine et al., 2021; Shweder et al., 1997), a strong conviction of the guilt of a defendant (called “moral certitude” in Roman Catholic canon law: Hahn, 2019), or the effect of violence on colonized people (which was known as “moral effect” among administrators of the British Empire, Elkins, 2022). Since we seek a technical definition, these omissions are by design. We sacrifice some overlap with non-technical usage to ensure that our definition picks out a class of phenomena that is unified enough to theorize about.

(b) Psychological: The definition should identify widely shared psychological phenomena.: A second requirement was that our definition should pick out psychological phenomena that are widely shared, so that the definition is applicable to multiple cultural groups. I noted earlier in the paper how most people view concerns with others’ welfare, rights, fairness, and justice as important and obligatory (Graham et al., 2013; Turiel, 2002). Naturally, people vary across situations and among themselves in how they map those concerns onto specific events and how they coordinate competing concerns mapped

onto the same situation. But this very variation in the application and balancing of moral concerns makes the definition amenable to cross-cultural research. The definition provides a framework for studying variations in a shared construct. Without the underlying shared construct, we would not know what to compare.

The components of the definition—concerns, judgments, reasoning, emotions, and actions—are all psychological phenomena that we can assess empirically. We can interview individuals about their judgments and reasoning about harmful actions, measure their facial or physiological reactions to harmful actions, or assess their actual or intended efforts to stop an ongoing harmful action. In some cases, it will be methodologically challenging to determine whether a person's judgment, reasoning, emotion, or action was in fact guided by moral concerns. This kind of indeterminacy is inherent to almost any psychological construct. There is no method that can tell with absolute certainty and in all situations whether a person is angry either. This methodological pickle calls for multiple methods—including behavioral experiments, naturalistic methods, physiological assessments, and interviews—as instantiated in recent multi-method work with both children and adults (Dahl, 2017a; Hepach et al., 2012; Langenhoff et al., in press; Tomasello, 2018).

(c) Empirical: The definition should incorporate characteristics that shape people's normative views.: The proposed definition identifies empirical, psychological phenomena involved in the formation of normative judgments, but it makes no commitment about which actions are right and which are wrong. Consider again the issue of abortion: According to the proposed definition, abortion is a moral issue both for those who are pro-life and those who are pro-choice (Dworkin, 1993; Gallup Inc., 2018; Goodwin & Landy, 2014). Pro-life individuals are concerned with the rights of the fetus, which they prioritize over women's rights over their bodies. Pro-choice individuals, though they also recognize the moral relevance of prenatal life, prioritize the rights of women to self-determination over the moral value of the fetus. Both perspectives fit within the moral domain as defined here, since they deal with obligatory concerns with rights and welfare. But the proposed definition does not commit the scientists studying pro-life and pro-choice views to treat one view as morally right and the other as morally wrong. Two researchers who personally disagree about women's right to choose can still work within the same scientific framework to study how people form moral judgments about abortion.

(d) Distinctive: The definition should separate moral judgments from other normative judgments.: By restricting moral concerns to obligatory concerns with others' welfare, rights, justice, and fairness, the proposed definition renders morality distinct from other normative considerations. First, it distinguishes judgments rooted in obligatory concerns, such as moral judgments, from judgments rooted in non-obligatory concerns, such as personal preferences (Dahl & Schmidt, 2018; Kant, 1785; Kohlberg, 1971). The obligatoriness of moral concerns is a formal feature of my definition. My definition of morality also has a substantive component. It defines moral concerns as obligatory concerns with others' welfare, rights, fairness, and justice. Moral judgments, reasoning, emotions, and actions are those that spring from such moral concern. This substantive definition thus sets

moral concerns apart from other obligatory concerns, such as conventional concerns with authorities or prudential concerns with one's own welfare.

Do Other Definitions of Morality Fit the Four Criteria?

My own definition of morality is one sensible definition of morality. Other definitions of morality can also meet the criteria of being technical, psychological, descriptive, and distinctive. To what extent have current theories provided definitions that meet the four criteria? I will consider three examples: Social Domain Theory, Theory of Dyadic Morality, and Moral Foundations Theory.¹²

Social Domain Theory—The definition closest to my proposed definition that of Social Domain Theory. Social Domain Theorists usually define morality as considerations and judgments about others' welfare, rights, justice, and fairness (Killen & Smetana, 2015; Turiel, 2015). In developing and defending his definition, Turiel has relied on the writings of philosophers such as Dworkin, Gewirth, Nussbaum, and Rawls. Turiel (2015) writes: "The substantive moral considerations [these philosophers] identified and analyzed include justice, rights, and civil liberties, with the promotion of human welfare and equal treatment as components" (pp. 1–2).

Turiel has been criticized for overstating the agreement among philosophers about what morality is (Machery & Stich, 2022; Stich, 2018). Even his critics were right, and I am not sure they are, Turiel never assumed that *all* philosophers, let alone all people, shared his definition of morality. He justified his definition on the grounds that it was shared by many philosophers, especially those in the rationalist tradition, and that it identified a distinctive psychological phenomenon. Turiel (1989) wrote that the definitions of the moral and conventional domains "were derived from a back and-forth process between the formulation of definitional parameters and data gathering" (p. 94). But even if his definition operates like a technical definition, Turiel has not usually called his definition technical—doing so might have spared him from some of the above criticisms.

We can also infer that Turiel's definition of morality is psychological, as it references psychological entities like considerations and judgments; that it is empirical, as it makes no assumptions about which judgments are morally right or wrong; and that it is distinctive, as it relies on concepts like welfare, rights, justice, and fairness.

This does not mean that every single writer within the Social Domain Theory tradition treat their definition of morality as technical, psychological, empirical, and substantive. When Nucci and Gingo (2011), writing within the tradition of Social Domain Theory, say that children distinguish "between morality (non-arbitrary and unavoidable features of social relations pertaining to matters of human welfare and fairness) and matters of convention (contextually dependent and agreed-upon social rules)" (p. 422), they do not mark their definition of morality as technical. Lacking such markings, some readers have come away thinking that Social Domain Theory lays claim to the One True Morality. Social Domain

¹²I thank a reviewer for suggesting that I include this discussion and another reviewer for prompting me to discuss Gray's Theory of Dyadic Morality.

theorists could reduce confusion, and help the field, by stating that their definition is a technical, psychological, empirical, and substantive one.

Theory of Dyadic Morality—A second common framework for psychological research on morality is the Theory of Dyadic Morality (Gray et al., 2012; Schein & Gray, 2018a, 2018b). As I mentioned earlier, this theory intimates several definitions of morality. At times, the theory suggests that all judgments of right and wrong are moral judgments, as when Schein and Gray (2018a) equate “moralization” with the processes by which “acts become wrong.” The Theory of Dyadic Morality has also argued that there are no perceived “harmless wrongs:” When someone judges an act to be wrong, they always perceive some possible or actual harm (Gray et al., 2014; Schein & Gray, 2018b). In the taxonomy of the present article, this represents the *normative* approach, as morality is taken to refer to all judgments of right and wrong.

Elsewhere, Schein and Gray (2018b) imply that moral norms comprise only a subset of norms about right and wrong actions (p. 36). The Theory of Dyadic Morality “suggests that harm separates ‘conventional’ negative norm violations from violations of morality” (Schein & Gray, 2018b, p. 60). This quote would indicate that Gray and colleagues take what I called a *distinctive* approach to defining morality—one that seeks to separate morality from at least some other normative considerations about right and wrong.

Although the Theory of Dyadic Morality hints at a distinctive approach to defining morality, it does not provide such a distinctive definition. The theory claims that perceived harm is a characteristic of most moral judgments, but presents this claim as an empirical hypothesis—not a definition. Schein and Gray (2018b) write that “the main prediction of dyadic morality [is] that acts are immoral to the extent that they are harmful” (p. 43) and then purport to provide evidence for the hypothesis. To test this prediction empirically, the theory must rely on a separate, non-harm criterion for identifying the set of moral judgments. On this set of moral judgments, the theory can test its prediction that all moral judgments involve perceived harm.

However, Schein and Gray (2018b) argue that a clear definition of morality is unattainable.¹³ They write that “any complex concept defies a complete philosophical definition” (p. 35) and morality is a “fuzzy psychological template” (p. 42). They write that “it is impossible to draw a firm in/out line around any rich human concept, as Wittgenstein (1953) discovered long ago when trying to determine the authoritative definition of ‘games’” (Schein & Gray, 2018b, p. 42). In their allusion to Wittgenstein, Schein and Gray appear to adopt what I have called a linguistic approach to defining morality. (Further suggesting a linguistic approach, some studies within the framework of the Theory of Dyadic Morality have assessed morality by asking participants how “immoral” an act was [Gray et al., 2022].) That is, they suggest

¹³Schein and Gray (2015b) do acknowledge Haidt’s definition of moral judgments as “evaluations (good vs. bad) of the actions or character of a person that are made with respect to a set of virtues held to be obligatory by a culture or subculture (Haidt, 2001, p. 817)” (Schein & Gray, 2018b, p. 35). Yet Haidt’s definition does not appear to circumscribe the morality that the Theory of Dyadic Morality aims to explain. Schein and Gray (2015b) refer to Haidt’s definition as “one popular working definition,” which is hardly a ringing endorsement in academes.

that a theory of morality should explain everything that people call “moral,” even if all the uses of “moral” cannot be captured by a single definition.

Wittgenstein did remind us that no single feature united all the things that people call “games.” Still, as I argued above, the polysemy and fuzziness of everyday language do not preclude scientists from adopting technical definitions. Rather, those very qualities of everyday language render technical definitions necessary for psychological science. (Indeed, Schein and Gray [2018b] embraced a technical redefinition of “harm.”) Accordingly, I believe that the Theory of Dyadic Morality would grow in clarity and testability if it committed to a technical—as well as a psychological, descriptive, and distinctive—definition of morality.¹⁴

Moral Foundations Theory—Moral Foundations Theory has alternated between normative and functionalist definitions of morality. As I have explained in earlier sections, both of these approaches run into difficulties.

Consider first the normative definition. In presenting Moral Foundations Theory, Graham et al. (2013) write that “the kinds of phenomena we’re studying” include “a common concern in third-party normative judgments” (p. 107). They further explain that “third-party moral judgments” are judgments that “condemn others for actions that have no direct consequences for the self” (p. 109). These statements fall under what I have called the “normative” approach to defining morality: Moral includes all normative judgments about right and wrong that people make from a third-party perspective. In earlier writings, Haidt and Joseph (2007) and Haidt and Graham (2007) similarly seemed to treat all normative judgments about right and wrong as falling within the moral realm. Readers may recall that, consistent with this normative definition, the Moral Foundations Questionnaire asks which factors people consider “when you decide whether something is right or wrong.”

As Graham and colleagues (2013) acknowledge, the broad, normative definition will oblige Moral Foundations Theory to incorporate additional foundations beyond the initial five (Care/harm, Fairness/cheating, Loyalty/betrayal, Authority/subversion, and Sanctity/degradation). They suggest that liberty/oppression (about political regimes) and efficiency/waste (about resource use) might be other foundations. But the number of foundations might keep growing. People make third-party judgments about right and wrong regarding all kinds of issues, some far removed from what most people would call moral issues. As Kohlberg

¹⁴What if the Theory of Dyadic Morality took first approach I mentioned in this section: the normative approach to defining morality? It would then have defined moral judgments as any judgment of right and wrong. In principle, the theory could then make the empirically testable hypothesis that all judgments of right and wrong involve perception of harm. Like all normative approaches to defining morality, this would strain the overlap between scientific and everyday usage of “moral.” Gray and his colleagues might accept this definitional strain, though. In some cases, they treat as “moral” judgments that fall outside the realm of what most people would call morality. According to the Theory of Dyadic Morality, even perceptions that agents are harming themselves, by doing something dangerous or unhealthy, can generate moral judgments (Schein & Gray, 2018b, p. 59). However, this normative definition of morality would expose the Theory of Dyadic Morality to a greater risk: unfalsifiability. The theory defines “harm” as virtually any possible or actual negative consequence to any entity subjectable to negative consequences, from deceased souls to the natural environment. In effect, the theory would be saying that people only judge an act as wrong if they thought the act might bring about something negative. If launched, this claim would land somewhere between the trivial and the tautologous. For how could anyone judge an act as wrong if they thoroughly knew that nothing bad could follow from it, not even a loss of self-respect? As Chomsky (1959) wrote in his review of Skinner’s *Verbal Behavior*, there is “no possible grounds for argument here” (p. 27)—not even for Kant (1785). For these reasons, I ultimately doubt that the Theory of Dyadic Morality would adopt the normative approach to defining morality.

(1971) noted, you can make a judgment about whether a person mixed a Martini in the right or wrong way. A theory that seeks to explain all normative judgments will also have to account for normative judgments about the mixing of Martinis.

Elsewhere, Moral Foundation Theorists have taken a functionalist approach to defining morality. Graham et al. (2011) borrowed a definition from Haidt and Kesebir (2010), which “defines moral systems by their *function*” (p. 368). The quoted passage from Haidt and Kesebir (2010) reads: “Moral systems are interlocking sets of values, virtues, norms, practices, identities, institutions, technologies, and evolved psychological mechanisms that work together to suppress or regulate selfishness and make social life possible” (p. 800). Also adopting Moral Foundations Theory, Hofman and colleagues (2018) describe morality as “a culturally transmitted set of normative values and rules that enable people to live together (more or less) in harmony” (p. 286). Here, too, morality is defined by its consequences: the enabling of peaceful coexistence.

I have already discussed the problems with using a functionalist definition of morality in psychological research. In short: We would be unable to tell whether some value or judgment qualifies as moral until the consequences of a particular value, habit, or judgment were known. Until we had resolved these complex, empirical questions about what enables peaceful coexistence, Moral Foundation Theory would not know which psychological features to place in the moral bucket.

By alternating between normative and functionalist definitions, Moral Foundations Theory has made it difficult to know which phenomena it treats as moral. Take the debate about whether authority concerns are “moral concerns.” If we adopt the normative definition of morality, virtually everyone—including Social Domain Theorists—would agree that concerns with authority commands fit the definition of morality. If we adopt the functionalist definition of morality, it will be hard to tell whether concerns with authority commands fit the definition of morality. Some authority concerns might fit (e.g., with the commands of Martin Luther King Jr.), others might not (e.g., with the commands of a nihilist or anarchist demagogue). And we cannot tell which authority concerns fit and which do not fit within the moral domain until we know their consequences.

There is a way to tweak Moral Foundations Theory to meet the criteria of a technical, psychological, empirical, and substantive definition of morality. Moral Foundation Theorists could stipulate that morality comprises the concerns that fall within the five articulated foundations (Care/harm, Fairness/cheating, Loyalty/betrayal, Authority/subversion, and Sanctity/degradation). This definition would not capture all judgments of right and wrong. There is also no guarantee that the modified definition would exclusively capture phenomena that enable peaceful coexistence. But this modified definition would promote rigorous inquiries and lucid debates. That, I believe, is the best we can aim for.

Will a technical, psychological, descriptive, and distinctive definition resolve the disagreements between Social Domain Theory, The Theory of Dyadic Morality, and Moral Foundations Theory? Will such a definition settle the questions about whether all moral judgments stem from perceptions of harm; whether morality is universal or culturally

variable; or whether morality is innate, learned, or constructed? A good and explicit definition is necessary for resolving these theoretical debates, not sufficient. This is as it should be. The job of a definition is to identify a set of empirical phenomena to investigate, not to make empirical claims about them. To resolve disagreements, we first need to identify the phenomena we want to make claims about. Next, we can study those phenomena to test our claims empirically. To test whether all swans are white, we need to know what a swan is. To test whether all whales descended from a four-legged land mammal, we need to know what we mean by “whale.” And to test whether moral judgments are rooted in perceptions of harm, or whether infants make moral judgments, we need to know what counts as a moral judgment. A suitable and explicit definition is the beginning of a resolvable inquiry into morality. Without a definition, there can be no resolution.

Conclusions

Morality is notoriously hard to define—so hard that many researchers have given up defining it. No matter the hardship, I have argued that psychologists need to define morality in order to advance our knowledge. Conversations without definitions are good for fiction, as in Raymond Carver’s story of love, but they are bad for science. This article considered what psychologists do when we define morality and why we need to do it. By getting clear on what we do when we define morality, it will be easier to devise suitable definitions.

I considered four approaches to defining morality. Each approach represented a different thing we might do when we define morality: the *linguistic* approach seeks to capture how people use the word “moral,” the *functionalist* seeks to capture morality by its function or consequences, the *evaluating* seeks to capture a set of morally good characteristics, and the *normative* seeks to capture all normative judgments. Each of these four definitional approaches encounters difficulties. But if we accept that psychological research on morality needs a definition of morality, it will not be enough to provide arguments *against* one definition; we will have to provide arguments *for* a better definition.

And so, to overcome the difficulties encountered by these four approaches, I proposed a definition that was technical, psychological, empirical, and distinctive. According to my proposed definition, morality comprises obligatory concerns with how to treat other sentient beings (i.e., with others’ welfare, rights, fairness, and justice), as well as the judgments, reasoning, emotions, and actions that spring from these concerns. This is not the only valid or sensible definition of morality for psychological research. I briefly discussed candidate definitions from Social Domain Theory, the Theory of Dyadic Morality, and Moral Foundations Theory, and noted ways in which they could amend their definitions.

How does having a definition help advance research on moral psychology? Two recent debates illustrate the benefits. One place where a definition comes in handy is in research on the emergence of morality in early childhood. In the mid-2000s, several pioneering studies were taken to support the claim that human infants are born with a moral sense (Bloom, 2013; Haidt & Kesebir, 2010; Hamlin et al., 2007; Hamlin, 2013). In these studies, infants in their first year watched a prosocial puppet help a neutral puppet reach some goal, for instance retrieving a ball, and an antisocial puppet hinder the neutral puppet from reaching

that goal. The basic finding was that infants were more likely to look or reach toward the prosocial puppet than toward the antisocial puppet (Hamlin et al., 2007; Hamlin & Wynn, 2011; Margoni & Surian, 2018). Many scholars took infants' preference for prosocial over antisocial puppets as evidence that infants make moral judgments (Hamlin, 2013).

Whether this finding demonstrates morality in infants depends on our definition of morality. If we define morality as a preference for helpful over hindering characters—setting aside whether that would be a good definition—then the findings would support the claim that infants have a moral sense. I proposed to define morality in terms of obligatory concerns, which guide judgments about right and wrong from a first-, second-, and third-person perspective. On this definition, infants have not yet acquired morality (Dahl, 2019). Infants' looking and reaching in these studies only reveal relative preferences (choosing one puppet over another), not categorical judgments (condemning the antisocial puppet, regardless of whom the antisocial puppet is compared to, Dahl et al., 2013). Moreover, even if infants prefer prosocial over antisocial others, there is no evidence that they negatively evaluate their own acts of hitting, biting, or kicking others—acts they engage in more often than older children and adults, often without provocation (Dahl, 2016). By my proposed definition, morality does not emerge until around the third birthday (Dahl, 2019; Tomasello, 2018). This illustrates that we can address the question of when morality emerges only after we have defined morality.

Another debate that demands a definition of morality is about whether the fundamental elements of morality are universally shared. Some have argued that the morality of one community differs fundamentally from the morality of another (Haidt & Graham, 2007; Levine et al., 2021; Machery, 2018; Shweder et al., 1997; Turiel, 2002). Others have proposed that the same fundamental moral concerns are evident in all communities (Turiel, 2002). But the question about universality, too, hinges on what we mean by morality. If by morality we mean all kinds of judgments about right and wrong, it is clear that morality in this sense varies from one community to another. In a religious community, the word of God is a source of judgments about right and wrong not shared by members of secular communities (Atran & Ginges, 2012; Srinivasan et al., 2019). But if we instead mean to ask whether people in all communities have obligatory concerns with others' welfare, rights, fairness, and justice—as I have proposed to do here—evidence compels us to conclude that morality, thus defined, *is* evident in all communities.

The truth of our claims always hinges on the meaning of our words. The truth of the claim that “our solar system has eight planets” depends what we mean by “planet.” By the standard definition of “planet,” adopted by the International Astronomical Union, the claim is true. By a revised definition of “planet” that also includes Pluto and the moons, the claim is false (Metzger et al., 2022). In the same way, the truth of claims like “infants have a moral sense” or “the fundamental elements of morality are universally shared” depends on the underlying definition of “morality.”

Approaching the end of this article, a reader might wonder if my proposal gets at the *true* or *real* morality. The reader might protest that I failed to identify a “natural kind” of morality (Stich, 2018). Perhaps the reader will wonder why they should accept my definition as the

final word when philosophers—in their centuries of unsurpassed definitional excellence—have offered conflicting conceptions of morality (Wiggins, 2006; Wynn & Bloom, 2014). But I do not claim to define the One True Morality. (Frankly, I would not know how to look for it.) And I never promised to find a natural kind. Instead, I proposed a definition that identifies a class of phenomena that merits psychological inquiry and invites psychological theorizing. I have also laid out four criteria against which other definitions of morality for psychological science can be measured.

It can feel disappointing to give up on the psychological study of the One True Morality. Disappointment, though, is not a sign of error. Disappointment is what we feel when our hopes have been dashed—here, our hopes for finding a definition of morality that is demonstrably better and truer than all others. Disappointment is also what we feel when we realize that our hopes were unrealistic.

To any psychologist who wishes to continue the quest for the One True Morality, we should ask: What reasons do we have for believing that there is One True Morality out there for psychologists to define? And how will we know when we have found it? Pending a successful quest for the ultimate definition, the following conclusions will pertain:

- Without definitions of morality, moral psychologists and their readers will not know what moral psychology seeks to explain.
- Some common approaches to defining morality—those I call linguistic, functionalist, evaluating, and normative—run into difficulties.
- To overcome these difficulties, we can articulate technical, psychological, empirical, and distinctive definitions of morality.
- I proposed one workable definition in this paper: obligatory concerns with how to treat other sentient beings (welfare, fairness, rights, and justice), as well as reasoning, judgments, emotions, and actions that spring from those concerns.

In the future, different camps may unite under a shared definition of morality. For now, the field can live with multiple definitions, provided we make those definitions explicit. Without explicit and workable definitions, we will—like Raymond Carver’s two couples—keep finding ourselves in impasses. With explicit and workable definitions, we can communicate better across research paradigms, separate definitional differences from empirical disagreements, and advance a cumulative psychological science of morality.

Acknowledgments

The writing of this manuscript was supported, in part, by a grant from the National Institute of Child Health and Human Development (R03HD087590). I thank Tal Waltzer, Margie Martinez, Charles Baxley, Mary Taylor Goeltz, and other members of the Developmental Moral Psychology Lab at UC Santa Cruz for their comments on earlier versions of this manuscript.

References

Alford CF (2002). *Whistleblowers: Broken lives and organizational power*. Cornell University Press.

- Altemeyer B, & Hunsberger B (2004). A revised Religious Fundamentalism Scale: The short and sweet of it. *International Journal for the Psychology of Religion*, 14(1), 47–54. 10.1207/s15327582ijpr1401_4
- Aquino K, & Reed A (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83(6), 1423–1440. 10.1037/0022-3514.83.6.1423 [PubMed: 12500822]
- Arsenio WF (2015). Moral psychological perspectives on distributive justice and societal inequalities. *Child Development Perspectives*, 9(2), Article 2. 10.1111/cdep.12115
- Asch S (1952). *Social psychology*. Prentice-Hall.
- Atran S, & Ginges J (2012). Religious and sacred imperatives in human conflict. *Science*, 336(6083), Article 6083. 10.1126/science.1216902
- Bandura A (1990). Selective activation and disengagement of moral control. *Journal of Social Issues*, 46(1), 27–46. 10.1111/j.1540-4560.1990.tb00270.x
- Bandura A (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, 31(2), 101–119. 10.1080/0305724022014322
- Bandura A (2016). *Moral disengagement: How people do harm and live with themselves*. Macmillan Higher Education.
- Bandura A (2018). A commentary on moral disengagement: The rhetoric and the reality. *The American Journal of Psychology*, 131(2), 246–251. 10.5406/amerjpsyc.131.2.0246
- Barrett LF., Lewis M., & Haviland-Jones JM. (Eds.). (2016). *Handbook of emotions* (4th ed.). The Guilford Press.
- Batson CD, Chao MC, & Givens JM (2009). Pursuing moral outrage: Anger at torture. *Journal of Experimental Social Psychology*, 45(1), 155–160. 10.1016/j.jesp.2008.07.017
- Bicchieri C (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.
- Blake PR, McAuliffe K, Corbit J, Callaghan TC, Barry O, Bowie A, Kleutsch L, Kramer KL, Ross E, Vongsachang H, Wrangham R, & Warneken F (2015). The ontogeny of fairness in seven societies. *Nature*, 528(7581), Article 7581. 10.1038/nature15703
- Bloom P (2013). *Just babies: The origins of good and evil*. Crown.
- Bredemeier BJ, & Shields DL (1986). Athletic aggression: An issue of contextual morality. *Sociology of Sport Journal*, 3(1), 15–28. 10.1123/ssj.3.1.15
- Browning CR (1992). *Ordinary men: Reserve police battalion 101 and the final solution in poland*. Harper Collins. <https://repository.library.georgetown.edu/handle/10822/851932>
- Campos JJ, Walle EA, Dahl A, & Main A (2011). Reconceptualizing emotion regulation. *Emotion Review*, 3(1), Article 1. 10.1177/1754073910380975
- Carpendale JIM, Hammond SI, & Atwood S (2013). A relational developmental systems approach to moral development. *Advances in Child Development and Behavior*, 45, 125–153. [PubMed: 23865115]
- Carver CS, & Scheier MF (2001). *On the self-regulation of behavior*. Cambridge University Press.
- Carver R (1981). *What we talk about when we talk about love*. Knopf Doubleday Publishing Group.
- Checkers. (2022). In Wikipedia. https://en.wikipedia.org/w/index.php?title=Checkers&oldid=1126223333#Invented_variants
- Chomsky N (1959). Review of Verbal behavior. By B. F. Skinner. *Language*, 35(1), 26–58. 10.2307/411334
- Cohen Priva U, & Austerweil JL (2015). Analyzing the history of Cognition using Topic Models. *Cognition*, 135, 4–9. 10.1016/j.cognition.2014.11.006 [PubMed: 25497481]
- Corzine A, & Dahl A (2022). Perceptions of interest and agency explain disagreements about sexual violations. *Journal of Social and Personal Relationships*, 026540752210743. 10.1177/02654075221074392
- Craig SC, Kane JG, & Martinez MD (2002). Sometimes you feel like a nut, sometimes you don't: Citizens' ambivalence about abortion. *Political Psychology*, 23(2), 285–301. 10.1111/0162-895X.00282
- Curry OS, Alfano M, Brandt MJ, & Pelican C (2021). Moral molecules: Morality as a combinatorial system. *Review of Philosophy and Psychology*. 10.1007/s13164-021-00540-x

- Cushman F, Gray K, Gaffey A, & Mendes WB (2012). Simulating murder: The aversion to harmful action. *Emotion*, 12(1), 2–7. 10.1037/a0025071 [PubMed: 21910540]
- Dahl A (2014). Definitions and developmental processes in research on infant morality. *Human Development*, 57(4), 241–249. 10.1159/000364919 [PubMed: 27570254]
- Dahl A (2016). Infants' unprovoked acts of force toward others. *Developmental Science*, 19(6), Article 6. 10.1111/desc.12342
- Dahl A (2017a). Ecological commitments: Why developmental science needs naturalistic methods. *Child Development Perspectives*.
- Dahl A (2017b). Two challenges for psychologists against, or in favor of, empathy. *Human Development*, 60, 193–200. 10.1159/000479845
- Dahl A (2019). The science of early moral development: On defining, constructing, and studying morality from birth. *Advances in Child Development and Behavior*, 56, 1–35. 10.1016/bs.acdb.2018.11.001 [PubMed: 30846044]
- Dahl A, Campos JJ, & Witherington DC (2011). Emotional action and communication in early moral development. *Emotion Review*, 3(2), 147–157. 10.1177/1754073910387948
- Dahl A, & Freda GF (2017). How young children come to view harming others as wrong: A developmental analysis. In Sommerville JA & Decety J (Eds.), *Social Cognition: Development across the life span* (pp. 151–184). Routledge.
- Dahl A, Gingo M, Uttich K, & Turiel E (2018). Moral reasoning about human welfare in adolescents and adults: Judging conflicts involving sacrificing and saving lives. *Monographs of the Society for Research in Child Development*, Serial No. 330, 83(3), 1–109. 10.1111/mono.12374
- Dahl A, Gross RL, & Siefert C (2020). Young children's judgments and reasoning about prosocial acts: Permission, obligation, and supererogation. *Cognitive Development*, 55, 100908. [PubMed: 32699466]
- Dahl A, & Killen M (2018a). Moral reasoning: Theory and research in developmental science. In Wixted JT & Ghetti S (Eds.), *The Stevens' handbook of experimental psychology and cognitive neuroscience*, 4th ed. (pp. 323–353). Wiley.
- Dahl A, & Killen M (2018b). Moral reasoning: Theory and research in developmental science. In Wixted JT & Ghetti S (Eds.), *The Stevens' handbook of experimental psychology and cognitive neuroscience*, 4th ed. (pp. 323–353). Wiley.
- Dahl A, & Kim L (2014). Why is it bad to make a mess? Preschoolers' conceptions of pragmatic norms. *Cognitive Development*, 32, 12–22. 10.1016/j.cogdev.2014.05.004 [PubMed: 28154468]
- Dahl A, & Schmidt MFH (2018). Preschoolers, but not adults, treat instrumental norms as categorical imperatives. *Journal of Experimental Child Psychology*, 165, 85–100. 10.1016/j.jecp.2017.07.015 [PubMed: 28826577]
- Dahl A, Schuck RK, & Campos JJ (2013). Do young toddlers act on their social preferences? *Developmental Psychology*, 49(10), Article 10. 10.1037/a0031460
- Dahl A, & Waltzer T (2018). Moral disengagement as a psychological construct. *American Journal of Psychology*, 131, 240–246.
- Dahl A, & Waltzer T (2020a). Constraints on conventions: Resolving two puzzles of conventionality. *Cognition*, 196, 104152. 10.1016/j.cognition.2019.104152 [PubMed: 31841815]
- Dahl A, & Waltzer T (2020b). Rationalization is rare, reasoning is pervasive. *Behavioral and Brain Sciences*, 43, e75. [PubMed: 32349809]
- Daston L (2022). *Rules: A short history of what we live by*. Princeton University Press.
- Davis DB (2006). *Inhuman bondage: The rise and fall of slavery in the new world*. Oxford University Press.
- Doris JM, & Murphy D (2007). From My Lai to Abu Ghraib: The Moral Psychology of Atrocity. *Midwest Studies In Philosophy*, 31(1), 25–55. 10.1111/j.1475-4975.2007.00149.x
- Dworkin R (1993). *Life's dominion: An argument about abortion and euthanasia*. Knopf.
- Elkins C (2022). *Legacy of violence: A history of the british empire*. Alfred A. Knopf.
- Ellemers N, van der Toorn J, Paunov Y, & van Leeuwen T (2019). The psychology of morality: A review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review*, 23(4), 332–366. 10.1177/1088868318811759 [PubMed: 30658545]

- Emerson MO, & Hartman D (2006). The rise of religious fundamentalism. *Annual Review of Sociology*, 32(1), 127–144. 10.1146/annurev.soc.32.061604.123141
- Fiske AP, & Rai TS (2014). *Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships*. Cambridge University Press.
- Friedman M (1962). *Capitalism and freedom*. University of Chicago Press.
- Frijda NH (1986). *The emotions*. Cambridge University Press.
- Frimer JA, Gaucher D, & Schaefer NK (2014). Political conservatives' affinity for obedience to authority is loyal, not blind. *Personality and Social Psychology Bulletin*, 40, 1205–1214. 10.1177/0146167214538672 [PubMed: 24972942]
- Gallup Inc. (2018, January 28). Abortion trends by party identification. Gallup.Com. <https://news.gallup.com/poll/246278/abortion-trends-party.aspx>
- Gelfand M (2018). *Rule makers, rule breakers: How tight and loose cultures wire our world*. Simon and Schuster.
- Gershoff ET (2002). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin*, 128(4), Article 4. 10.1037//0033-2909.128.4.539
- Gert B, & Gert J (2017). The definition of morality. In Zalta EN (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017, Fall 2017). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2017/entries/morality-definition/>
- Goodwin GP, & Landy JF (2014). Valuing different human lives. *Journal of Experimental Psychology: General*, 143, 778–803. 10.1037/a0032796 [PubMed: 23647311]
- Graham J, Haidt J, Koleva S, Motyl M, Iyer R, Wojcik SP, & Ditto PH (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, 47, 55–130. 10.1016/B978-0-12-407236-7.00002-4
- Graham J, & Iyer R (2012). The unbearable vagueness of “essence”: Forty-four clarification questions for Gray, Young, and Waytz. *Psychological Inquiry*, 23(2), 162–165. 10.1080/1047840X.2012.667767
- Graham J, Nosek BA, & Haidt J (2012). The moral stereotypes of liberals and conservatives: Exaggeration of differences across the political spectrum. *PLOS ONE*, 7(12), e50092. 10.1371/journal.pone.0050092 [PubMed: 23251357]
- Graham J, Nosek BA, Haidt J, Iyer R, Koleva S, & Ditto PH (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2), 366–385. 10.1037/a0021847 [PubMed: 21244182]
- Gray K., & Graham J. (Eds.). (2018). *Atlas of moral psychology*. Guilford Publications.
- Gray K, MacCormack JK, Link to external site, this link will open in a new window, Henry T, Link to external site, this link will open in a new window, Banks E, Link to external site, this link will open in a new window, Schein C, Armstrong-Carter E, Link to external site, this link will open in a new window, Abrams S, Link to external site, this link will open in a new window, Muscatell KA, & Link to external site, this link will open in a new window. (2022). The affective harm account (AHA) of moral judgment: Reconciling cognition and affect, dyadic morality and disgust, harm and purity. *Journal of Personality and Social Psychology*, 123(6), 1199–1222. 10.1037/pspa0000310 [PubMed: 35357881]
- Gray K, Schein C, & Ward AF (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, 143(4), 1600–1615. 10.1037/a0036149 [PubMed: 24635184]
- Gray K, Young L, & Waytz A (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23(2), Article 2. 10.1080/1047840X.2012.651387
- Greene JD (2007). The biology of morality: Neuroscientists respond to Killen and Smetana. *Human Development*. http://www.karger.com/ProdukteDB/Katalogteile/issn/_0018_716X/HDE-Letters-to-Editor-10-09-2007.pdf
- Greene JD (2008). The secret joke of Kant's soul. In Sinnott-Armstrong W (Ed.), *Moral psychology, Vol 3: The neuroscience of morality: Emotion, brain disorders, and development* (pp. 35–80). MIT Press.
- Greene JD (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin Press.

- Greene JD (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics*, 124(4), 695–726. 10.1086/675875
- Greene JD, Sommerville RB, Nystrom LE, Darley JM, & Cohen JD (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108. 10.1126/science.1062872 [PubMed: 11557895]
- Grossmann I, Weststrate NM, Ardelt M, Brienza JP, Dong M, Ferrari M, Fournier MA, Hu CS, Nusbaum HC, & Vervaeke J (2020). The science of wisdom in a polarized world: Knowns and unknowns. *Psychological Inquiry*, 31(2), 103–133. 10.1080/1047840X.2020.1750917
- Hahn J (2019). Moral certitude: Merits and demerits of the standard of proof applied in roman catholic jurisprudence. *Oxford Journal of Law and Religion*, 8(2), 300–325. 10.1093/ojlr/rwz012
- Haidt J (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834. 10.1037/0033-295X.108.4.814 [PubMed: 11699120]
- Haidt J (2008). Morality. *Perspectives on Psychological Science*, 3(1), 65–72. [PubMed: 26158671]
- Haidt J (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.
- Haidt J (2013). Moral psychology for the twenty-first century. *Journal of Moral Education*, 42(3), Article 3. 10.1080/03057240.2013.817327
- Haidt J, Bjorlund F, & Murphy S (2000). Moral dumbfounding: When intuition finds no reason [Unpublished manuscript].
- Haidt J, & Graham J (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20(1), 98–116. 10.1007/s11211-007-0034-z
- Haidt J, & Kesebir S (2010). Morality. In Fiske ST, Gilbert DT, & Lindzey G (Eds.), *Handbook of social psychology*, 5th edition (pp. 797–832). Wiley.
- Hamlin JK (2013). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science*, 22(3), 186–193. 10.1177/0963721412470687
- Hamlin JK, & Wynn K (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, 26(1), Article 1. 10.1016/j.cogdev.2010.09.001
- Hamlin JK, Wynn K, & Bloom P (2007). Social evaluation by preverbal infants. *Nature*, 450(7169), 557–559. 10.1038/nature06288 [PubMed: 18033298]
- Hardecker S, Schmidt MFH, & Tomasello M (2017). Children’s developing understanding of the conventionality of rules. *Journal of Cognition and Development*, 18(2), Article 2. 10.1080/15248372.2016.1255624
- Helwig CC (2006). Rights, civil liberties, and democracy across cultures. In Killen M & Smetana JG (Eds.), *Handbook of moral development* (pp. 185–210). Lawrence Erlbaum Associates Publishers.
- Henrich J, Heine SJ, & Norenzayan A (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. 10.1017/S0140525X0999152X [PubMed: 20550733]
- Henrich J, & Henrich N (2010). The evolution of cultural adaptations: Fijian food taboos protect against dangerous marine toxins. *Proceedings of the Royal Society B: Biological Sciences*, 277(1701), Article 1701. 10.1098/rspb.2010.1191
- Hepach R, Vaish A, & Tomasello M (2012). Young children are intrinsically motivated to see others helped. *Psychological Science*, 23(9), 967–972. 10.1177/0956797612440571 [PubMed: 22851443]
- Hertz SG, & Krettenauer T (2016). Does moral identity effectively predict moral behavior? A meta-analysis. *Review of General Psychology*, 20(2), 129–140. 10.1037/gpr0000062
- Heyes C (in press). Rethinking norm psychology. *Perspectives on Psychological Science*. <https://ora.ox.ac.uk/objects/uuid:55723002-f750-4fc4-b24a-baf3e3fc3c4f>
- Hoffman ML (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press.
- Hofmann W, Meindl P, Mooijman M, & Graham J (2018). Morality and self-control: How they are intertwined and where they differ. *Current Directions in Psychological Science*, 27(4), 286–291. 10.1177/0963721418759317

- Holyoak KJ, & Powell D (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin*, 142(11), 1179–1203. 10.1037/bul0000075 [PubMed: 27709981]
- Hume D (1738). *A treatise of human nature*. Oxford University Press.
- Hunt L (2008). *Inventing human rights: A history*. W. W. Norton & Company.
- Hutcherson CA, & Gross JJ (2011). The moral emotions: A social–functionalist account of anger, disgust, and contempt. *Journal of Personality and Social Psychology*, 100(4), Article 4. 10.1037/a0022408
- Jacobson D (2012). Moral dumbfounding and moral stupefaction. In Timmons M (Ed.), *Oxford studies in normative ethics* (Vol. 2, pp. 289–316). Oxford University Press.
- Jennings PL, Mitchell MS, & Hannah ST (2015). The moral self: A review and integration of the literature. *Journal of Organizational Behavior*, 36(S1), S104–S168. 10.1002/job.1919
- Jost JT, & Kay AC (2010). Social justice: History, theory, and research. In Fiske ST, Gilbert DT, & Lindzey G (Eds.), *Handbook of social psychology*, Vol. 2, 5th ed (pp. 1122–1165). John Wiley & Sons Inc.
- Kahn PH (1992). Children’s obligatory and discretionary moral judgments. *Child Development*, 63(2), 416–430. 10.2307/1131489 [PubMed: 1611944]
- Kant I (1785). *Groundwork of the metaphysics of morals* (Paton HJ, Trans.). Routledge.
- Kelly D, Stich S, Haley KJ, Eng SJ, & Fessler DM (2007). Harm, affect, and the moral/conventional distinction. *Mind & Language*, 22(2), Article 2.
- Killen M (2016). Morality: Cooperation is fundamental but it is not enough to ensure the fair treatment of others. *Human Development*, 59(5), 324–337. 10.1159/000454897
- Killen M, & Dahl A (2020). The moral obligations of conflict and resistance. *Behavioral and Brain Sciences*, 43, e75. 10.1017/S0140525X19002401 [PubMed: 32349809]
- Killen M, & Dahl A (2021). Moral reasoning enables developmental and societal change. *Perspectives on Psychological Science*, 16, 1209–1225. 10.1177/1745691620964076 [PubMed: 33621472]
- Killen M, Elenbaas L, & Rizzo MT (2018). Young children’s ability to recognize and challenge unfair treatment of others in group contexts. *Human Development*, 61(4–5), Article 4–5. 10.1159/000492804
- Killen M, & Smetana JG (2015). Origins and development of morality. In Lamb M (Ed.), *Handbook of child psychology and developmental science* (7th ed., Vol. 3, pp. 701–749). Wiley-Blackwell.
- Killen M, & Turiel E (1998). Adolescents’ and young adults’ evaluations of helping and sacrificing for others. *Journal of Research on Adolescence*, 8(3), 355–375. 10.1207/s15327795jra0803_4
- Knobe J (2003). Intentional action and side effects in ordinary language. *Analysis*, 63(3), 190–194.
- Kochanska G, & Aksan N (2006). Children’s conscience and self-regulation. *Journal of Personality*, 74(6), 1587–1618. 10.1111/j.1467-6494.2006.00421.x [PubMed: 17083659]
- Kohlberg L (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In Mischel T (Ed.), *Psychology and genetic epistemology* (pp. 151–235). Academic Press.
- Korsgaard CM (1996). *The sources of normativity*. Cambridge University Press.
- Köymen B, Lieven E, Engemann DA, Rakoczy H, Warneken F, & Tomasello M (2014). Children’s norm enforcement in their interactions with peers. *Child Development*, 85(3), Article 3. 10.1111/cdev.12178
- Krebs DL (2008). Morality: An evolutionary account. *Perspectives on Psychological Science*, 3, 149–172. [PubMed: 26158933]
- Krettenauer T, & Hertz S (2015). What develops in moral identities? A critical review. *Human Development*, 58(3), Article 3. 10.1159/000433502
- Kuhn TS (1962). *The structure of scientific revolutions*. University of Chicago Press.
- Langenhoff AF, Dahl A, & Srinivasan M (in press). Preschoolers learn new moral and conventional norms from direct experiences. *Journal of Experimental Child Psychology*, 20.
- Laupa M (1991). Children’s reasoning about three authority attributes: Adult status, knowledge, and social position. *Developmental Psychology*, 27(2), Article 2.
- Lazarus RS (1991). *Emotion and adaptation*. Oxford University Press.

- Lazarus RS (2001). Relational meaning and discrete emotions. In Scherer KR, Schorr A, & Johnstone T (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 37–67). Oxford University Press.
- Legros S, & Cislighi B (2020). Mapping the social-norms literature: An overview of reviews. *Perspectives on Psychological Science*, 15(1), 62–80. 10.1177/1745691619866455 [PubMed: 31697614]
- Levine S, Kleiman-Weiner M, Schulz L, Tenenbaum J, & Cushman F (2020). The logic of universalization guides moral judgment. *Proceedings of the National Academy of Sciences*, 117(42), 26158–26169. 10.1073/pnas.2014505117
- Levine S, Rottman J, Davis T, O'Neill E, Stich S, & Machery E (2021). Religious affiliation and conceptions of the moral domain. *Social Cognition*, 39(1), 139–165. 10.1521/soco.2021.39.1.139
- Lewis D (1969). *Convention: A philosophical study*. Harvard University Press.
- Lewis M (2007). Self-conscious emotional development. In Tracy JL, Robins RW, & Tangney JP (Eds.), *The self-conscious emotions: Theory and research* (pp. 134–149). Guilford Press.
- Machery E (2018). Morality: A historical invention. In Gray K & Graham J (Eds.), *The atlas of moral psychology* (pp. 259–265). Guilford Press.
- Machery E, & Stich S (2022). The moral/conventional distinction. In Zalta EN (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2022). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2022/entries/moral-conventional/>
- Malle BF (2021). Moral judgments. *Annual Review of Psychology*, 72, 293–318. 10.1146/annurev-psych-072220-104358
- Margoni F, & Surian L (2018). Infants' evaluation of prosocial and antisocial agents: A meta-analysis. *Developmental Psychology*, 54(8), Article 8. 10.1037/dev0000538
- McAlister AL, Bandura A, & Owen SV (2006). Mechanisms of Moral Disengagement in Support of Military Force: The Impact of Sept. 11. *Journal of Social and Clinical Psychology*, 25(2), Article 2. 10.1521/jsep.2006.25.2.141
- McNamara P (2011). Supererogation, inside and out: Toward an adequate scheme for common sense morality. In Timmons M (Ed.), *Oxford studies in normative ethics: Vol. Volume 1* (pp. 202–235). Oxford University Press.
- Melville H (1851). *Moby-Dick, or the whale*. Wordsworth Editions.
- Metzger PT, Grundy WM, Sykes MV, Stern A, Bell JF, Detelich CE, Runyon K, & Summers M (2022). Moons are planets: Scientific usefulness versus cultural teleology in the taxonomy of planetary science. *Icarus*, 374, 114768. 10.1016/j.icarus.2021.114768
- Mill JS (1879). *Utilitarianism*. Longmans, Green and Company.
- Miller D (2017). Justice. In Zalta EN (Ed.), *The Stanford Encyclopedia of Philosophy* (2017th ed.). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2017/entries/justice/>
- Miller JG, Bersoff DM, & Harwood RL (1990). Perceptions of social responsibilities in India and in the United States: Moral imperatives or personal decisions? *Journal of Personality and Social Psychology*, 58(1), 33–47. [PubMed: 2308074]
- Miller JG, Goyal N, & Wice M (2017). A cultural psychology of agency: Morality, motivation, and reciprocity. *Perspectives on Psychological Science*, 12(5), 867–875. 10.1177/1745691617706099 [PubMed: 28972843]
- Moore GE (1903). *Principia ethica*. Cambridge University Press.
- Moors A, & Scherer KR (2013). The role of appraisal in emotion. In Robinson MD, Watkins ER, & Harmon-Jones E (Eds.), *Handbook of cognition and emotion* (pp. 131–155). Guilford Press.
- Mukherjee S (2016). *The gene: An intimate history*. Simon and Schuster.
- Mulligan K, & Scherer KR (2012). Toward a working definition of emotion. *Emotion Review*, 4(4), 345–357.
- Murray JAH, & Bradley H (1900). *A new English dictionary on historical principles: Founded mainly on the materials collected by the Philological Society, Volume IV* Clarendon Press. <https://archive.org/details/oed04arch/page/n1/mode/2up>
- Nagel T (1989). *The view from nowhere*. Oxford University Press.

- Nucci LP (1981). Conceptions of personal issues: A domain distinct from moral or societal concepts. *Child Development*, 52(1), 114–121. 10.2307/1129220
- Nucci LP (2001). *Education in the moral domain*. Cambridge University Press.
- Nucci LP (2004). The promise and limitations of the moral self construct. In *Changing conceptions of psychological life* (pp. 49–70). Lawrence Erlbaum Associates Publishers.
- Nucci LP (2019). Character: A developmental system. *Child Development Perspectives*, 13(2), 73–78. 10.1111/cdep.12313
- Nucci LP, & Gingo M (2011). The development of moral reasoning. In Goswami U (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (Version 2nd, 2nd ed., pp. 420–445). Wiley-Blackwell.
- Nucci LP, Guerra N, & Lee J (1991). Adolescent judgments of the personal, prudential, and normative aspects of drug usage. *Developmental Psychology*, 27(5), Article 5.
- Nucci LP, Narvaez D, & Krettenauer T (2014). *Handbook of moral and character education* (2nd ed.). Routledge.
- Nucci LP, & Turiel E (1993). God's word, religious rules, and their relation to Christian and Jewish children's concepts of morality. *Child Development*, 64(5), 1475. 10.2307/1131547
- Nucci LP, Turiel E, & Roded AD (2017). Continuities and discontinuities in the development of moral judgments. *Human Development*, 60, 279–341. 10.1159/000484067
- Oakley B., Knafo A., Madhavan G., & Wilson DS. (Eds.). (2011). *Pathological altruism*. Oxford University Press.
- Osiel MJ (2001). *Obeying orders: Atrocity, military discipline & the law of war*. Routledge. 10.4324/9781315125442
- Padilla-Walker LM., & Carlo G. (Eds.). (2016). *Prosocial development: A multidimensional approach*. Oxford University Press.
- Petruzzello M (2016). Is a tomato a fruit or a vegetable? Britannica. <https://www.britannica.com/story/is-a-tomato-a-fruit-or-a-vegetable>
- Piaget J (1932). *The moral judgment of the child*. Free Press.
- Rawls J (1971). *A theory of justice*. Harvard University Press.
- Rhodes M, & Chalik L (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science*, 24(6), Article 6. 10.1177/0956797612466267
- Rogoff B (2003). *The cultural nature of human development*. Oxford University Press.
- Romero A (2012). When whales became mammals: The scientific journey of cetaceans from fish to mammals in the history of science. In Romero A(Ed.), *New approaches to the study of marine mammals*. IntechOpen. 10.5772/50811
- Royzman EB, & Borislow SH (2022). The puzzle of wrongless harms: Some potential concerns for dyadic morality and related accounts. *Cognition*, 220, 104980. 10.1016/j.cognition.2021.104980 [PubMed: 34990961]
- Royzman EB, Kim K, & Leeman RF (2015). The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. *Judgment & Decision Making*, 10(4), Article 4.
- Royzman EB, Leeman RF, & Baron J (2009). Unsentimental ethics: Towards a content-specific account of the moral-conventional distinction. *Cognition*, 112(1), 159–174. 10.1016/j.cognition.2009.04.004 [PubMed: 19439283]
- Scanlon TM (1998). *What we owe to each other*. Harvard University Press.
- Schein C, & Gray K (2018a). Moralization: How acts become wrong. In *Atlas of moral psychology* (pp. 363–370). The Guilford Press.
- Schein C, & Gray K (2018b). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, 22(1), 32–70. 10.1177/1088868317698288 [PubMed: 28504021]
- Schmidt MFH, & Tomasello M (2012). Young children enforce social norms. *Current Directions in Psychological Science*, 21(4), Article 4. 10.1177/0963721412448659
- Schwartz SH, & Bardi A (2001). Value hierarchies across cultures: Taking a similarities perspective. *Journal of Cross-Cultural Psychology*, 32(3), Article 3. 10.1177/0022022101032003002

- 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. 10.1080/09515089.2012.727135
- Sharp FC (1896). The limitations of the introspective method in ethics. *The Philosophical Review*, 5(3), 278–291.
- Shweder RA, Mahapatra M, & Miller JG (1987). Culture and moral development. In Kagan J & Lamb S (Eds.), *The emergence of morality in young children* (pp. 1–83). University of Chicago Press.
- Shweder RA, Much NC, Mahapatra M, & Park L (1997). The “big three” of morality (autonomy, community, divinity) and the “big three” explanations of suffering. In Brandt AM & Rozin P (Eds.), *Morality and health* (pp. 119–169). Taylor & Frances/Routledge.
- Silk JB, & House BR (2011). Evolutionary foundations of human prosocial sentiments. *Proceedings of the National Academy of Sciences*, 108(Supplement_2), 10910–10917. 10.1073/pnas.1100305108
- Simpson J., & Weiner E. (Eds.). (1989). *The Oxford English dictionary (Second Edition)*. Oxford University Press.
- Sinnott-Armstrong W, & Wheatley T (2012). The disunity of morality and why it matters to philosophy. *The Monist*, 95(3), 355–377. 10.5840/monist201295319
- Skitka LJ, Hanson BE, Morgan GS, & Wisneski DC (2021). The psychology of moral conviction. *Annual Review of Psychology*, 72, 347–366. 10.1146/annurev-psych-063020-030612
- Smedslund J (1991). The pseudoempirical in psychology and the case for psychologic. *Psychological Inquiry*, 2(4), 325–338. 10.1207/s15327965pli0204_1
- Smetana JG (2013). Moral development: The social domain theory view. *Oxford Handbook of Developmental Psychology*, 1, 832–866.
- Srinivasan M, Kaplan E, & Dahl A (2019). Reasoning about the scope of religious norms: Evidence from Hindu and Muslim children in India. *Child Development*, 90, e783–e802. 10.1111/cdev.13102 [PubMed: 29896806]
- Sripada CS, & Stich S (2006). A framework for the psychology of norms. In Carruthers P, Laurence S, & Stich S (Eds.), *The innate mind Volume 2: Culture and cognition* (pp. 280–301). Oxford University Press.
- Stanford PK (2018). The difference between ice cream and Nazis: Moral externalization and the evolution of human cooperation. *Behavioral and Brain Sciences*, 41, 1–13. 10.1017/S0140525X17001911
- Stich S (2018). The quest for the boundaries of morality. In *The Routledge handbook of moral epistemology* (pp. 15–37). Taylor & Francis. <https://www.researchwithnj.com/en/publications/the-quest-for-the-boundaries-of-morality>
- Thompson RA (2012). Whither the preconventional child? Toward a life-span moral development theory. *Child Development Perspectives*, 423–429. 10.1111/j.1750-8606.2012.00245.x
- Tisak MS (1993). Preschool children’s judgments of moral and personal events involving physical harm and property damage. *Merrill-Palmer Quarterly*, 375–390.
- Tisak MS, & Turiel E (1984). Children’s conceptions of moral and prudential rules. *Child Development*, 55(3), 1030–1039. 10.2307/1130154
- Tomasello M (2016). *A natural history of human morality*. Harvard University Press.
- Tomasello M (2018). The normative turn in early moral development. *Human Development*, 61, 248–263.
- Triandis HC (2001). Individualism-collectivism and personality. *Journal of Personality*, 69(6), Article 6. 10.1111/1467-6494.696169
- Turiel E (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
- Turiel E (1989). Domain-specific social judgments and domain ambiguities. *Merrill-Palmer Quarterly* (1982-), 35, 89–114.
- Turiel E (2002). *The culture of morality: Social development, context, and conflict*. Cambridge University Press.

- Turiel E (2010). The relevance of moral epistemology and psychology for neuroscience. In Zelazo PD, Chandler M, & Crone E (Eds.), *Developmental social cognitive neuroscience* (pp. 313–331). Taylor & Francis.
- Turiel E (2015). Moral development. In Lerner RM (Ed.), *Handbook of child psychology and developmental science* (Vol. 1, pp. 484–522). John Wiley & Sons, Inc.
- Turiel E (2018). Reasoning at the root of morality. In Gray K & Graham J (Eds.), *The atlas of moral psychology* (pp. 9–19). Guilford Press.
- Turiel E, Chung E, & Carr JA (2016). Struggles for equal rights and social justice as unrepresented and represented in psychological research. *Advances in Child Development and Behavior*, 50, 1–29. 10.1016/bs.acdb.2015.11.004 [PubMed: 26956068]
- Turiel E, & Dahl A (2019). The development of domains of moral and conventional norms, coordination in decision-making, and the implications of social opposition. In Bayertz K & Boughley N(Eds.), *The normative animal: On the biological significance of social, moral, and linguistic norms* (pp. 195–213). Oxford University Press.
- Turiel E, Dahl A, & Besirevic Z (2018). Thoughts, emotions, and sentiments in the development of justice. In LeBar M (Ed.), *Becoming just* (pp. 119–150). Oxford University Press.
- Turiel E, Hildebrandt C, & Wainryb C (1991). Judging social issues: Difficulties, inconsistencies, and consistencies. *Monographs of the Society for Research in Child Development*, 56(2), Article 2. 10.2307/1166056
- Turiel E, Killen M, & Helwig CC (1987). Morality: Its structure, functions, and vagaries. In Kagan J & Lamb S (Eds.), *The emergence of morality in young children* (pp. 155–243). University of Chicago Press.
- Vicente A, & Falkum IL (2017). Polysemy. In *Oxford Research Encyclopedia of Linguistics*. 10.1093/acrefore/9780199384655.013.325
- Wade N (2004, July 30). Francis Crick, co-discoverer of DNA, dies at 88. *The New York Times*. <https://www.nytimes.com/2004/07/30/us/francis-crick-co-discoverer-of-dna-dies-at-88.html>
- Wainryb C (1991). Understanding differences in moral judgments: The role of informational assumptions. *Child Development*, 62(4), Article 4. 10.2307/1131181
- Wainryb C (2011). ‘And so they ordered me to kill a person’: Conceptualizing the impacts of child soldiering on the development of moral agency. *Human Development*, 54(5), 273–300. 10.1159/000331482
- Wainryb C, Brehl BA, & Matwin S (2005). Being hurt and hurting others: Children’s narrative accounts and moral judgments of their own interpersonal conflicts. *Monographs of the Society for Research in Child Development*, 70(3), Article 3. 10.1111/j.1540-5834.2005.00350.x
- Wakslak CJ, Jost JT, Tyler TR, & Chen ES (2007). Moral outrage mediates the dampening effect of system justification on support for redistributive social policies. *Psychological Science*, 18(3), 267–274. 10.1111/j.1467-9280.2007.01887.x [PubMed: 17444925]
- Walker LJ (2014). Moral personality, motivation, and identity. In Killen M & Smetana JG (Eds.), *Handbook of moral development*, 2nd ed (2nd ed., pp. 497–519). Psychology Press.
- Waytz A, Dungan J, & Young L (2013). The whistleblower’s dilemma and the fairness–loyalty tradeoff. *Journal of Experimental Social Psychology*, 49(6), Article 6. 10.1016/j.jesp.2013.07.002
- Wierzbicka A (2007). “Moral sense.” *Journal of Social, Evolutionary, and Cultural Psychology*, 1(3), 66–85. 10.1037/h0099824
- Wiggins D (2006). *Ethics: Twelve lectures on the philosophy of morality*. Harvard University Press.
- Williams B (1973). A critique of utilitarianism. In Smart JJC & Williams B (Eds.), *Utilitarianism: For and against* (pp. 77–149). Cambridge University Press.
- Wittgenstein L (1953). *Philosophical investigations* (Anscombe GE, Trans.). Basil Blackwell.
- Wynn K, & Bloom P (2014). The moral baby. In Killen M & Smetana JG(Eds.), *Handbook of moral development* (2nd ed., pp. 435–453). Psychology Press.
- Yilmaz O, Harma M, Bahçekapili HG, & Cesur S (2016). Validation of the moral foundations questionnaire in Turkey and its relation to cultural schemas of individualism and collectivism. *Personality and Individual Differences*, 99, 149–154. 10.1016/j.paid.2016.04.090

Zhang Y, & Li S (2015). Two measures for cross-cultural research on morality: Comparison and revision. *Psychological Reports*, 117(1), 144–166. 10.2466/08.07.PR0.117c15z5 [PubMed: 26241096]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

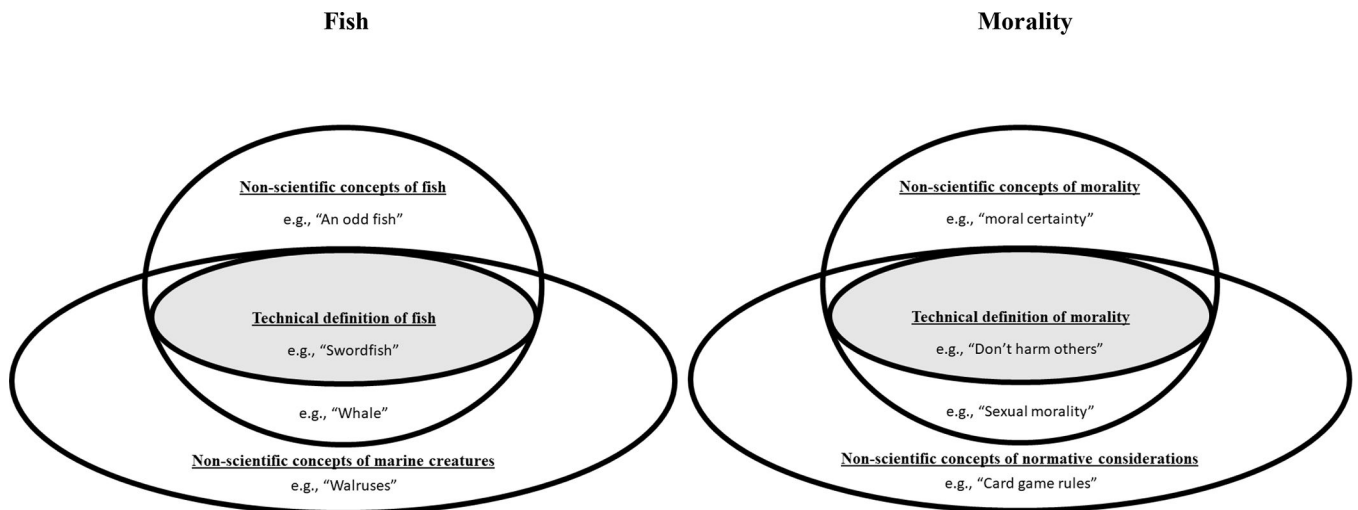


Figure 1: Representation of Relation between Non-Scientific Concepts and Technical Definitions for “Fish” and “Morality”.

Note. The quoted phrases are examples of what would fit into each segment of the diagram. The particular visualization of a technical definition of morality illustrates the definition I propose in this paper. For other technical definitions, the everyday phrases included and excluded in the definition will be different.

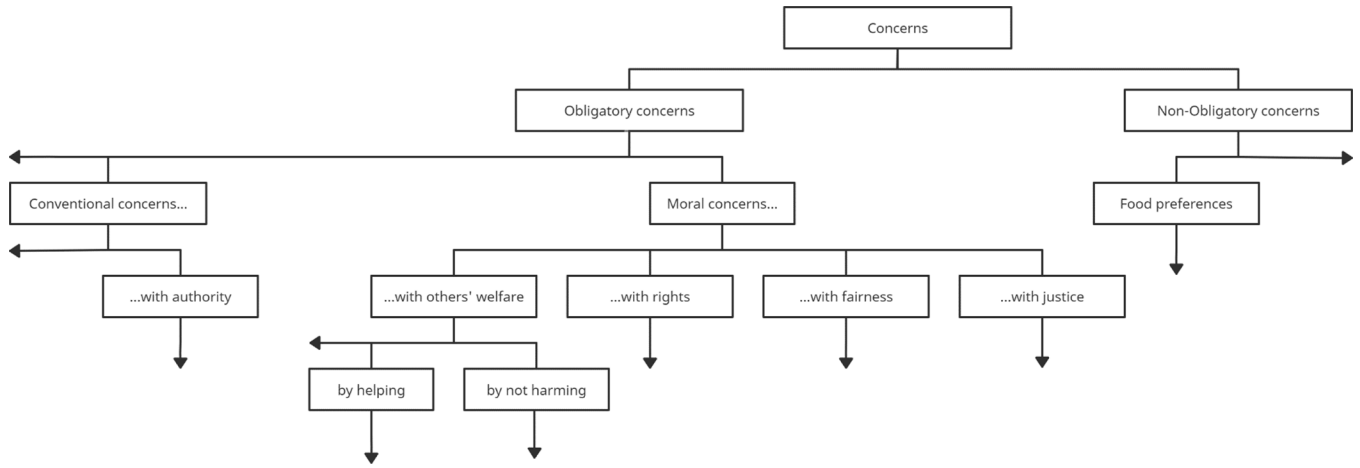


Figure 2: Classification of Concerns.

Note. The figure illustrates different kinds of concerns. The top-level division separates obligatory and non-obligatory concerns. (To reiterate: A person deems a concern obligatory if they deem it wrong to lack this concern.) Obligatory concerns are divided into moral, conventional, and other obligatory concerns. Moral concerns are subdivided into concerns with others’ welfare, rights, fairness, and justice. Vertical arrows indicate further concerns within a given category. For instance, concerns with rights may be divided into concerns with property rights, rights to free speech, and so on (Helwig, 2006; Hunt, 2008; Turiel et al., 2016). Horizontal arrows indicate additional categories at the same level. For instance, in addition to moral and conventional concerns, individuals may also deem prudential (personal wellbeing) or pragmatic (material order) concerns obligatory (Dahl & Kim, 2014; Smetana, 2013; Srinivasan et al., 2019). Non-obligatory concerns are as multifarious as—if not more multifarious than—obligatory ones. Beyond food preferences, there are preferences for sports teams, romantic partners, and so on, many of which we might call “values” (Schwartz & Bardi, 2001)

Table 1

Overview of Approaches to Defining Morality

| Common approaches to definitions | Difficulty | To overcome the difficulty, a definition has to... | Proposed definition |
|---|---|--|--|
| (a) Linguistic <i>Capturing how people use the word "moral"</i> | "Moral" has many uses in ordinary language | ...overlap with some but not all common uses of the word "moral" (technical definition) | Morality is... |
| (b) Functionalist <i>Capturing morality by its function or consequences</i> | Psychological phenomena do not have a 1:1 mapping with functions or consequences | ... identify widely shared psychological phenomena (psychological definition) | ... rooted in a set of obligatory concerns |
| (c) Evaluating <i>Capturing the set of (morally) right characteristics</i> | An empirical science has no methods for separating right from wrong actions or traits | ...that shape people's normative judgments without committing the researchers to specific views about right and wrong actions or traits (empirical definition) | ...that give rise to reasoning, judgments, emotions, and actions about right and wrong |
| (d) Normative <i>Capturing all considerations about right and wrong</i> | Normative considerations comprise a heterogeneous group of social and non-social considerations | ...and separate moral judgments from other normative judgments. (distinctive definition) | ...ways of treating others (i.e., concerns with welfare, rights, fairness, and justice). |