

What Are the Irreducible Basic Elements of Morality? A Critique of the Debate Over Monism and Pluralism in Moral Psychology

Bree Beal 

Department of English, Clemson University

Abstract

The debate between monists and pluralists in moral psychology has been framed as an argument over the number of “irreducible basic elements” that can be used to describe the extent of the moral domain: Do all moral values ultimately reduce to one principle (i.e., monism), or are there multiple irreducibly distinct moral values (i.e., pluralism)? I critique the premise of this debate, arguing that the breadth of the moral domain cannot be adequately represented, understood, or explained in terms of moral values. Instead, an adequate account of moral psychology must explain moral phenomena in terms of more basic elements: ontological frames.

Keywords

monism vs. pluralism debate, values, morality, moral psychology, moral cognition, ontological framing

I do not regard the Nazis, as some people do, as literally pathological or insane, only as wickedly wrong, totally misguided about the facts, for example in believing that some beings are subhuman, or that race is central, or that Nordic races alone are truly creative, and so forth. I see how, with enough false education, enough widespread illusion and error, men can, while remaining men, believe this and commit the most unspeakable crimes.

—Isaiah Berlin (2001, p. 15)

In 2013, Graham et al. published an essay developing the pluralistic model of moral psychology promulgated by Haidt (2012) and defending his moral foundations theory (MFT) against monists who would try to reduce the complexity of the moral world to the action of a single moral value. Graham et al. (2013) chose as an epigraph for their article the following quote from Isaiah Berlin, in which the philosopher expressed his own commitment to value pluralism:

I came to the conclusion that there is a plurality of ideals, as there is a plurality of cultures and of temperaments. . . . There is not an infinity of

[values]: the number of human values, of values which I can pursue while maintaining my human semblance, my human character, is finite—let us say 74, or perhaps 122, or 27, but finite, whatever it may be. (Berlin, 2001, p. 12)

I remember agreeing wholeheartedly with Graham et al.’s argument that the moral world is irreducibly complex and embracing their critique of monistic theories that would reduce all moral judgments to a consideration of the value of justice or the principle of doing no harm. I have not changed my mind on these points, but in the years since I first read Graham et al. (2013), I began to question whether value pluralism is adequate for explaining the complexity and dynamism of everyday moral cognition. There seem to be many interpersonal and intergroup moral differences, as well as dynamic fluctuations of moral judgment within individuals, that cannot be explained in terms of values at

Corresponding Author:

Bree Beal, Clemson University, Department of English, 815 Strode Tower, Clemson, SC 29631

E-mail: bbeal@clemson.edu

all. Indeed, for evidence of such, we need look no further than Berlin (2001).

Let us examine the quote I have chosen for my epigraph. What kind of theory underlies Berlin's statement about the Nazis? Should it be called *value pluralism*? Here, Berlin does not speak of "values" but instead of "the facts." Nazis commit "unspeakable crimes" because of their misguided beliefs "that some beings are subhuman, or that race is central, or that Nordic races alone are truly creative, and so forth." Are such beliefs values? Are facts, whether right or "wickedly wrong," aptly called values? It seems to me that such beliefs are something else altogether and that it is important to appreciate what they are and how they contribute to morality.

The word *value* is polysemous (i.e., has a variety of related meanings; see D'Andrade, 2008). For the current discussion, however, we need concern ourselves only with two of these possible meanings, which are often conflated by moral psychologists: values as principles and value that is attributed to particular things. The first of these refers to the principles one lives by. As such, values apply across multiple contexts and inform our treatment of multiple things. For example, to hold the value of justice is to believe that one ought to conform to the principle of justice across varied relationships and contexts. On the other hand, to say that "Nordic races alone are truly creative" is not to express a generalizable principle for moral living but to make an ontological claim about a specific group. It is true that the ontological claim above involves *evaluation* (i.e., "truly creative"), but such an evaluative belief differs from the traditional concept of a moral value in its concreteness: An ontological claim is always *of* or *about* a specific thing (whether that thing is an individual or a group, a relationship or a way of life, a concrete object or an abstract idea), whereas principles are abstracted away from particular relationships. Thus, although it is possible to ontologically define a particular entity in terms of the quality of justice—as in the statement "you are just"—the attempt to define someone or something as a principle—e.g., "you are justice"—is either absurd or poetic. Such considerations clearly distinguish the value we attribute to particular things in defining what or who they are from a value in the sense of an abstract principle that guides our behavior in diverse contexts and relationships.

These considerations also show that Berlin's description of the relation between ontological beliefs (e.g., "the facts") and moral judgments and behaviors (e.g., the Nazis' ideology and their "unspeakable crimes") is not a theory of value pluralism but something else—something that does not yet have a name. One might put the primary claim of this nameless theory thus: One's evaluative framing of reality (e.g., "the facts") is

the primordial determiner of one's moral judgments, and this influence of what we might call *ontological framing* upon our moral sensibility is, to a significant degree, independent of our values. One can perceive this primordiality and independence of ontological framing by analyzing Berlin's account of the Nazis. He does not claim that the Nazis had different values than others—they were not so "pathological" or "insane" as to reject the principle of justice, for instance. Rather, as Berlin suggests, the Nazis used their "wickedly wrong" ontological claims precisely in the attempt to *justify* their actions. Thus, although I would not claim that Nazi values were the same in every way as those of their victims, I would agree with the suggestion that the primary moral difference between Nazis and their victims is not in their abstract values so much as their specific ontological commitments.

The concept I am using to explicate Berlin's nameless theory is (ironically) partly inspired by Heidegger's account of ontological understanding. Heidegger, who would later join the Nazi party, pointed out in his opus *Being and Time* (1927/2010) that our ontological beliefs emerge only once we are already in the world—desiring, fearing, pursuing projects that matter to us. Our ontologies are thus organized in advance by our nature as caring beings concerned with our existence. It is not the case that entities exist objectively and are subsequently discovered and interpreted through the categories of pure reason. Rather, the meaning of an object or person or way of life emerges out of our particular way of navigating our situation, and we only objectify and measure and reason about the objects that emerge for us after we have already framed these objects in relation to our existence, our cares, our projects in the world.

Consequently, primordial understanding first emerges at an implicit level, as we more or less skillfully navigate our environment. And because this primordial understanding emerges out of our attractions and fears, pursuits and failures, our definitions of things tend to be evaluative or normative. For instance, if I am trying to achieve a particular goal, I am likely to understand whatever helps me achieve the goal as being, in a certain sense, good. This evaluative resonance is part of my ontological framing of that entity, implicit in identifiers such as *teammate*, *helper*, *compatriot*, and so on, whereas other ontological frames carry an opposite resonance (e.g., *enemy*, *hinderer*, *obstacle*, *traitor*). As I will show, this normative structure of ontological framing is the primordial origin and primary organizing principle of moral cognition.

I make two key claims in this essay. One is that moral cognition, morality, and the moral domain are determined by ontological framing before and, to a significant

degree, independently of the influence of abstract moral values. This is the claim I see as implicit in Berlin's statement above. Like Berlin, I do not argue that values do not exist or that they do not matter, but I do seek to place discussions of moral values within their proper place. Moral psychologists often treat the concept of the moral value as a fundamental unit of explanation, failing to realize that a moral value can emerge only once we already have developed a moral sensibility through the action of ontological framing. This is my second claim: Treating the moral value as the fundamental unit of morality is a significant limitation of current theories.

In this essay, I focus my critique on the debate to which I alluded in my first paragraph. Graham et al. (2013) frame their dispute with monists as an argument over the number of "irreducible basic elements" that can be used to "represent, understand, and explain the breadth of the moral domain" (p. 56). Do all moral values ultimately reduce to one principle (i.e., monism), or are there multiple irreducibly distinct moral values (i.e., pluralism)? This debate illustrates precisely the problem with many theories in moral psychology: Theorists assume that the "irreducible basic elements" of the moral domain are moral values and then imagine that they can explain the dynamics of moral cognition in terms of these elements. However, regardless of whether we decide that there are one, several, or many irreducibly distinct values, no explanation that takes moral values as the primordial unit of the moral domain will come anywhere close to explaining the dynamism of everyday moral cognition. For this, we must ground our theories in an analysis of ontological framing.

I contextualize the critique that follows with an overview of the current debate. MFT pluralists have aggressively attacked at least two kinds of moral monism: classic Kohlbergian monism, in which moral cognition is defined as "justice reasoning" (Kohlberg, 1984, pp. 215–217), and, more recently, the monism proposed in the theory of dyadic morality (TDM), in which moral cognition is said to be all about dyadic "harm" (Gray, 2014; Gray & Keeney, 2015a, 2015b; Gray, Schein, & Ward, 2014; Gray, Young, & Waytz, 2012; Royzman, Kim, & Leeman, 2015; Schein & Gray, 2015, 2018). Against these monistic hypotheses, MFT pluralists argue that, in addition to harm and justice considerations, relevant moral values include considerations of loyalty, authority, purity, and liberty (Graham, 2015; Graham et al., 2011, 2013, 2018; Haidt, 2012; Haidt & Bjorklund, 2008; Haidt, Graham, & Ditto, 2015; Rottman, Keleman, & Young, 2014). In recent years, this argument has been waged largely through empirical studies, with MFT pluralists showing that harm considerations are not the only relevant factor in determining participants' sense of moral violation and TDM monists arguing that when

studies are properly done we can see that morality is all about real or imagined harm. Although it seems to me that the pluralists win this argument—morality is about more than just harm—their framing of the stakes of the debate is misleading, and their explanation for why we have multiple moral values or value domains is far from adequate for explaining the real diversity of the moral domain and the real dynamism of moral cognition.

The following critique exposes why the moral domain cannot be adequately described in terms of moral values, whether one, several, or many. The most fundamental dimension of my argument is interpretive or philosophical: I show that the current debate over whether the moral domain is best described using one or several values is predicated upon flawed understandings of what morality is and how it emerges in cognition. This fundamental dimension of my argument is also inductive: I investigate an established empirical finding to identify the essential features of moral cognition, using these features to develop a model of moral psychology that is, in key respects, entirely new. Finally, this dimension is also *reductive* in the sense that I expose a unit of analysis that is more basic than the moral value and propose that we can explain moral phenomena in terms of the ontological frame—the real irreducible basic element of morality. A second dimension of my argument is practical, showing how ontological framing allows us to explain current findings and predict future ones much better than theories that explain moral phenomena primarily in terms of moral values. The philosophical and practical dimensions of this essay thus expand into a new model of moral psychology: one that explains moral cognition in terms of multifarious and dynamic ontological relationships.

The critique that follows is focused, as mentioned above, on one particular iteration of the debate over monism and pluralism that has gotten much attention of late. For principled reasons, I neglect older iterations of this debate. For instance, I completely ignore the metaphysical debate that goes by the same name. The reason for this exclusion of moral metaphysics should be clear by the end of the essay; however, for now I will offer as a preliminary explanation two observations about metaphysical definitions of morality. First, moral metaphysicians seek to identify the basic elements of a normative (i.e., rational, valid, or good) moral system, whereas I argue it is important for moral psychologists to appreciate the dark sides of morality (e.g., the ways in which morality functions to promote war, genocide, slavery, and the like). Dark sides are categorically excluded from a domain that is defined by light. Thus, the question of what makes up the moral domain will have totally different answers depending on if one is

asking the question in a normative metaphysical sense or in a descriptive psychological sense. Second, metaphysical definitions of morality are static (i.e., right and wrong are what they are and do not change), whereas I argue that a model of moral psychology must be sensitive to the dynamism of moral cognition in the everyday, and I point out that this sensitivity is possible only when one defines morality nonmetaphysically. (Metaphysical pluralists might point to the conflict between competing values or value domains; such conflict, however, does not express a dynamism of the moral domain itself but merely a limited kind of moral complexity.) In short, I ignore the longstanding debate over monism and pluralism among philosophers not because I think it is unimportant but rather because I think it is a categorically different debate—albeit a debate that ought to be informed by the descriptive model I propose in this essay.

I also neglect the earlier moral psychology debate between Turiel and Shweder in which the contemporary argument between TDM monists and MFT pluralists has its roots (see, e.g., Shweder, 1991; Shweder, Mahapatra, & Miller, 1987; Turiel, Killen, & Helwig, 1987). Although there are striking continuities between these two iterations of the debate over monism and pluralism, there are enough differences that each may need to be treated separately. For instance, it seems to me that Shweder's version of pluralism is much more amenable to the model I propose than the pluralism expressed in MFT. Thus, although this essay is critical of MFT pluralism, I also point to harmonies between Shwederian pluralism and my model. I am not arguing against pluralism *per se* but rather against the presuppositions that inform the current debate over monism and pluralism and that pervade the current field of moral psychology.

Monists and Pluralists Define Morality

TDM monists and MFT pluralists ground their positions in claims about the nature of morality that are distinct but that share a similar limitation. Although exposing this limitation is not the stated goal of this essay, it provides a useful contrast to the definition I offer in the next section—a definition I consider to be crucial given that a proper understanding of the primordial source of morality and moral cognition enables us to see the essential problem with describing moral cognition in terms of moral values. Thus, I begin by critiquing the respective definitions of morality accepted by TDM monists and MFT pluralists.

MFT pluralists define the moral domain functionally in terms of what morality does.¹ Over the course of evolutionary time, they argue, morality has promoted

cooperation; thus, MFT pluralists propose that a species of moral value called a moral foundation is an evolved predisposition that tends to promote cooperation in human groups (Graham et al., 2013; Haidt, 2012; Haidt & Kesebir, 2010). However, one might raise a cluster of related objections against the proposal that conduciveness to cooperation is a useful criterion for distinguishing moral values from nonmoral values and, more broadly, for defining the moral domain. First, it would seem obvious that many nonmoral features also promote cooperation, such as the tendency to imitate others (e.g., Henrich & Gil-White, 2001), the capacity for language, and even self-interest when cooperation is profitable (e.g., Smith, 1776/1904; Trivers, 1971). Moreover, it is equally clear that morality sometimes impedes cooperation—consider the moral conflicts on display within and between societies all over the world. Indeed, the very theories that posit that morality functions to promote cooperation also suggest that this in-group cooperation promotes intergroup competition (e.g., Darwin, 1871; Greene, 2013; Haidt, 2012; Sober & Wilson, 1998). Finally, when we consider the issue from a proximate psychological level of analysis (see, e.g., Mayr, 1961; Tinbergen, 1963), there is no good reason to deny that moral behaviors might at times promote neither cooperation nor competition (e.g., private pieties). But to define morality by its various functions as something that promotes either cooperation or competition—and sometimes both and other times neither—while acknowledging that many patently nonmoral features do the same is effectively to offer no definition at all. Thus, *function* does not help us to distinguish the moral domain from the nonmoral domain, and the fundamental question of how to distinguish moral values from other kinds of values cannot be addressed by a functionalist approach to defining morality.

TDM monists take a different approach to answering the question of what morality is. Gray et al. (2012) claim that “the essence of morality is mind perception.” What they mean is that “all moral judgments can be explained by appealing to a dyadic template: two perceived minds—a moral agent characterized by agency and a moral patient characterized by experience” (p. 103). Gray et al.'s central analogy for essence is Picasso's distillation of the essence of a bull, wherein the artist produces a series of progressively minimalist drawings until the final image is a stick figure of an archetypal bull. I think they make a convincing case that a dyadic relation between two persons is the central or archetypal form of moral relationship.² However, this characterization is essentially different from the common meaning of essence as a thing's indispensable quality or qualities—that without which a being is no longer itself—see, for example, the first definition of *essence*

(n.d.) in Merriam-Webster's online dictionary. Gray et al. also invoke the more common definition with their claim that a dyadic template explains "*all* [emphasis added] moral judgments." And, problematically, Gray et al. then equivocate on these two meanings of essence when they try to support their central claim that mind perception is the essence of morality. Although they offer evidence that a dyadic relation between a moral agent and suffering patient is a minimalist representation of an archetypal form of moral relationship—like Picasso's essence of morality—they then try to leverage this evidence to support the much stronger claim that this moral dyad can explain all moral judgments.

This much stronger claim that a perceived interpersonal dyad is the essence of morality in the sense of an indispensable quality is no more supported by their argument than the nonexistence of hornless bulls is proved by Picasso's portraits. But the weaker claim that the moral dyad is a minimalist representation of a common or archetypal moral relation that "fits the majority of moral situations" (Gray et al., 2012, p. 103) does not speak to the more fundamental and significant question of what morality's indispensable qualities are. Indeed, the vast majority of moral psychologists have always assumed, without having good support for this assumption, that morality is ultimately all about interpersonal relations (e.g., Gilligan, 1982; Kohlberg, 1971, 1981; Piaget, 1948/1965; Turiel, 1983). However, we must acknowledge exceptions to this common assumption that morality is essentially interpersonal and social. People may feel moral responsibilities toward a great variety of things, including animals, plants, spirits, ecosystems, our planet, our universe, ideas and ideals, priceless works of art, familiar places, sacred or sentimental objects, treasured ways of life, and so on. At least some of these things are not perceived as possessing minds, even by those who feel morally responsible for them. It is thus misleading to claim that the essence of morality—in the sense of an indispensable quality that can explain all moral judgments—is mind perception. Acknowledging exceptions to the dyadic template means acknowledging that we must dig deeper to discover what I call the proper essence of morality: the properties shared by all of our various moral judgments and intuitions.

More recently, TDM monists have adopted a slightly different approach to defining morality. Schein and Gray (2018) distill morality down to three components: "norm violations, negative affect, and dyadic harm" (p. 4). Although these components sometimes attend a moral judgment, I note that none of them is essential or necessary to morality. Norm violations come the closest to being an essential component, but even this condition must be nuanced, because, as Gray et al.

(2012) acknowledge, "moral judgments also apply to good deeds" (p. 3). In my own definition of the essence of morality, I therefore make the more precise claim that an essential quality of morality is that it is normative, in both positive (e.g., good deeds) and negative (e.g., norm violations) senses. However, we must also acknowledge that, because normativity is a broader category than morality, the quality of normativity is necessary but not sufficient for defining the moral domain.

Do Schein and Gray's other two criteria lead to a sufficient definition? In a word, no. The criteria of negative affect and dyadic harm are both nonessential to morality. Take negative affect: Not only can one feel positive affect in making a moral judgment, there is no good reason to deny that people may often perform moral computations in the absence of any affect at all. People with trait alexithymia, for instance, whose emotional processing is disrupted, do not lack the ability to make moral judgments. Their judgments tend to be utilitarian; they rely more on rational predictions of the outcomes of their moral choices than on strong feelings of empathy (Patil & Silani, 2014). Thus, their moral cognition is affected by their diminished affect, but this does not mean it is any less a form of moral cognition than that of neurotypical populations (see also May, 2018). Moreover, even if affects were an essential component of morality, we must acknowledge that they are not specific to morality because we also feel affects outside of the moral domain. So we are still lacking any criterion that could distinguish the moral domain categorically from the nonmoral domain. Is this criterion harm? Once again, the answer is no. MFT pluralists make a strong case that moral judgments are not inherently tied to harm because the world abounds with cases of harmless wrongs (Graham, 2015; Graham et al., 2011, 2013, 2018; Haidt, 2012; Haidt et al., 2015; Haidt, Koller, & Dias, 1993) and wrongless harms (Kelly, Stich, Haley, Eng, & Fessler, 2007). Thus, like Gray et al.'s distillation of the essence of morality, Schein and Gray's account provides a characterization of a prototypical moral interaction, but we must look elsewhere for the proper essence of morality.

In sum, the above analyses show that definitions of morality from TDM monists and MFT pluralists are both limited in the specific sense that they try to define morality in terms of a particular form or outcome of morality, substituting this for the more general definition that we need. For instance, if we rigorously pursue the task of defining morality in terms of its function, we discover such a multiplicity of functions that we lose any definite border separating the moral from the nonmoral—and the discussion above shows that no account of morality's evolutionary origins, however accurate, can justify the choice of cooperation as the

sole defining function of morality. Likewise, when we try to define morality like TDM monists—whether in terms of the archetype of an interpersonal dyad or as an emergent feature of norm violations, negative affect, and dyadic harm—we impose a narrow border around the moral domain that reveals nothing about the many variants that do not fit the prototype.

The limitations of both approaches to defining morality are a problem for the debate over monism and pluralism as it is currently framed. Most obviously, without a definitive way of distinguishing morality from nonmorality, we cannot argue with much precision over how many moral values make up the moral domain. However, more important than the precision with which this debate is argued is the fact that the entire argument is falsely premised: A question such as “How many moral values can explain the breadth of the moral domain?” presumes the breadth of the moral domain is explicable in terms of moral values. And this is simply not the case. Thus, we can see precisely the problem with the premise of this current debate between monists and pluralists when we appreciate that morality and the moral domain emerge from something more fundamental than moral values or, indeed, any particular form or function of morality. Thus, I offer a rigorous definition of morality, arguing that this definition suggests a need to dig down below the level of moral values to trace the emergence of moral cognition in ontological relationships.

The Essence of Morality

I do not mean to suggest with the above critique that psychologists know nothing about the moral domain. Even without a theoretically rigorous definition of morality, it is possible to empirically “operationalize” morality, and there is one widely accepted operationalization that might provide a starting point for our search. Turiel (1983) defined the moral domain with several criteria. His innovative experiments with children revealed that, across diverse societies, moral rules are characteristically understood as being more serious, universal, punishable, and context-independent than nonmoral conventions. This distinction between moral rules and social conventions has been observed in both adults and children in many studies (for reviews, see Nucci, 2001; Smetana, 2006; Smetana, Jambon, & Ball, 2014; Tisak, 1995; Turiel, 1998, 2002, 2006). And although the specific content of moral rules and values varies by culture (e.g., Graham et al., 2013; Haidt, 2012; Shweder et al., 1987; Shweder, Much, Mahapatra, & Park, 1997), the formal distinction between moral rules and conventions has been observed in both Western and non-Western societies, including India (Bersoff &

Miller, 1993; Madden, 1992; Miller & Bersoff, 1992), South Korea (Song, Smetana, & Kim, 1987), Hong Kong (Yau & Smetana, 2003), Indonesia (Carey & Ford, 1983), Nigeria (Hollos, Leis, & Turiel, 1986), Zambia (Zimba, 1987), Brazil (Nucci, Camino, & Milnitsky-Sapiro, 1996), Colombia (Ardila-Rey & Killen, 2001), and the Virgin Islands (Nucci, Turiel, & Encarnacion-Gawrych, 1983).

Given this extensive empirical support, it seems plausible that Turiel’s criteria are roughly accurate. However, this claim must be understood with the proper nuance. Turiel (1983) has also proposed that moral rules can be defined by their thematic content: Specifically, moral rules are concerned with justice, rights, and welfare. I have already discussed some of the substantial evidence against this claim, which includes evidence that there are harms that some cultures or individuals do not consider wrong (e.g., Kelly et al., 2007), as well as wrongs that do not involve harm, justice, or individual rights (e.g., Graham, 2015; Graham et al., 2011, 2013, 2018; Haidt, 2012; Haidt et al., 1993; 2015). I am convinced by these critiques of what I call Turiel’s thematic content claim. Yet we must be careful not to overextend the criticism, throwing the baby of Turiel’s formal distinction out with the bathwater of his content claim.

The hypothesis that moral rules are seen as more serious, universal, punishable, and context-independent than conventions is independent of the hypothesis that moral rules are exclusively concerned with justice, rights, and welfare. And although there is good evidence against the latter content claim, I am not aware of any strong evidence that would repudiate the formal distinction between moral rules and conventions. Indeed, the pluralistic argument that there are moral values or principles other than harm, justice, and rights, which bear a moral signature, presumes the validity of some formal distinction between moral and nonmoral principles; otherwise, it would not make any sense to speak of moral values or principles at all. Thus, although I agree with the pluralists and reject Turiel’s claim that moral cognition is essentially concerned with welfare, justice, and rights, I nevertheless accept Turiel’s and others’ suggestion that humans naturally distinguish between moral content and nonmoral content from an early age. And I also accept the evidence suggesting that Turiel’s criteria of seriousness, universality, punishability, and context independence are a good place to start for understanding the difference between moral and nonmoral norms and defining the moral domain.

Turiel’s criteria are a good place to start, but they are by no means adequate for defining the essence of morality. After all, one could follow a moral rule accidentally or mindlessly, giving the appearance of morality

without actually engaging in moral cognition. Or one may treat a rule as being serious, universal, punishable, and context-independent simply because one has been told so—similar to what Piaget (1948/1965) described as moral “heteronomy”—in which case one is demonstrating an understanding of language and social norms but not a cognitive capacity that is distinctively moral. Consider that we can program self-driving cars to follow rules we have designated as moral and to navigate issues of justice, harm, welfare, rights, and so on—yet it would be absurd to study these cars to learn about moral cognition. These considerations show that the difference between a genuine exercise of moral cognition and a heteronomous performance of morality is not in the objective content of the rules that are followed but rather in how the rules are understood, whether this understanding is explicit or intuitive. In short, moral cognition must emerge from a moral interpretation of things. So, to define the essence of morality, we must elucidate the basic conditions of such a moral interpretation: the criteria that can distinguish, for instance, incidental adherence to moral rules from the genuine practice of morality. We can pursue these criteria by considering the qualities that are necessary for interpreting moral rules as context-independent, universal, and so on. Although Turiel does not identify these qualities, his criteria can be seen as a set of useful symptoms that we can investigate deductively to discover morality’s essential features.

What do I mean it is possible to deduce the essence of morality? Philosophers might want to suggest I am seeking a “transcendental” definition of morality—something that, famously, has been done before (e.g., Kant, 1785/2002). However, when moral metaphysicians define morality’s transcendental conditions, they are trying to identify the conditions for any normatively valid moral claim or action. Moral psychologists need to identify something more basic: the essential conditions for any form of morality whatsoever—including whatever moral forms we ourselves consider invalid, irrational, or immoral. Only in this way can we adequately characterize the moral cognition of Nazis, to take one example. An astute reader might perceive that I am describing the difference between a prescriptive theory of moral metaphysics and a descriptive theory of moral psychology. Unfortunately, moral psychologists have often equivocated between psychological and metaphysical ways of defining morality, and we have yet to identify, in the broadly descriptive sense necessary for a psychological theory, what the essence of morality is. I acknowledge psychologists’ historical hesitance regarding such a project (e.g., Schein & Gray, 2018), a hesitance that is grounded in philosophical skepticism about the possibility of essentialist definitions (e.g.,

Wittgenstein, 1953/1997). But I simply echo the skeptic’s challenge back to the skeptics: “Look and see whether there is anything common to all” (Wittgenstein, 1953/1997, p. 31) forms of morality—that is, all moral judgments, theories, intuitions, and so on. I pursue my essentialist definition using Turiel’s “symptoms.” Theorists who doubt the validity of my definition are invited to provide counterexamples—forms of morality that do not involve the criteria below.

In defining the essential conditions of morality, I deductively interrogate Turiel’s symptoms by asking the following question: What are the necessary conditions of understanding oneself or others to be bound by a universally binding, context independent, serious, and punishable obligation? We can deduce at least three conditions:

1. Normativity: If morality involves obligation, it is a form of normativity. That is, it involves seeing things as good and bad, right and wrong, and so on. However, morality is not the only form of normativity. For instance, we can speak of a right and a wrong way to shoot a basketball. And nonmoral conventions are also normative. Thus, although a capacity to interpret the world normatively is an essential condition for the emergence of morality, this condition is not specific to morality, and so it is not sufficient for defining the moral domain.
2. Agency: A moral obligation to act in a certain way implies an ability to do so—or, at least, an ability to try, even if one cannot succeed. So, to hold ourselves or others morally responsible, we must attribute agency to ourselves and/or others. As with normativity, however, the condition of agency attribution is not specific to morality. An obligation to obey a nonmoral convention implies exactly the same attribution of agency. So we must answer another question: What is it that distinguishes moral from conventional norms?
3. Inherent value: A universally valid and context-independent obligation to protect or respect another being implies that that being is inherently (or intrinsically—see discussion below) worthy of respect or protection. That is, per Turiel’s conditions, moral obligations are not understood as being grounded in subjective opinion (i.e., they are universal) or external conditions (i.e., they are context-independent); thus, the only remaining basis for a moral obligation is the quality or qualities attributed to the thing itself. This grounding of the experience of moral responsibility in attribution of what we might

call *inherent value* makes intuitive sense. It would seem absurd to say that something or someone ought to be universally treated with respect, ought to be preserved or protected, ought not to be harmed, ought to be treated justly, and so on, while also claiming that nothing and no one is worthy of respect, protection, care, justice, and so on. Indeed, where such inherent value is not perceived or attributed, then values and rules are seen as situational, relative to context and one's practical goals. One ought to turn a screw clockwise if, and only if, one wants to screw it in, and social conventions are similarly conditional. Thus, we can deduce from Turiel's symptoms the fact that an attribution of worthiness, or inherent value, is the most distinctive and essential condition for the emergence of morality.³ (So much the better if reflection reveals that every actual theory of moral metaphysics really is grounded in implicit or explicit claims about what precisely is valuable in some universal and context-independent sense.)

I offer a few related caveats. First, we must respect the distinction between the metaphysical claim that humans really have agency or freedom of will and the psychological claim I am making here—that in feeling or imposing a sense of moral obligation, we implicitly or explicitly attribute agency to ourselves and/or others. I am neither claiming nor denying that people have free will. I am just pointing out that we implicitly attribute agency when we hold someone morally responsible. Likewise, I am not claiming that certain beings really have inherent value—that would be a metaphysical claim. Instead, I am saying only that some attribution of inherent value is folded into the idea of a universally valid and context-independent obligation. The ability/tendency to attribute agency and inherent value are thus essential features of moral normativity, without which moral cognition would not be what it is.

My next caveat is about my terminology. Is *inherent value* the right term? Should our third criterion instead be called “intrinsic goodness” (Moore, 1903), “final value” (Korsgaard, 1983), “primary good” (Rawls, 1971), or “good in itself” (Kant, 1785/2002)? I chose *value* instead of *goodness* because the attribution of both positive value (e.g., goodness) and negative value (e.g., evil) can have moral implications. As for the word *inherent*, I am less committed: It may be that *value in itself* or *intrinsic value* are more accurate terms. However, such fine distinctions may be precisely missing the point. Debates among philosophers over whether the morally

relevant form of value is an object's noninstrumental final or inherent value or, more narrowly, intrinsic value that an entity gives to itself may matter for moral metaphysics (see, e.g., Kant, 1785/2002; Korsgaard, 1983; Lewis, 1946). But these distinctions would be appropriate for moral psychology only if it were shown that in everyday moral cognition people really do make them, and these distinctions really are decisive for moral judgment. For instance, if it were shown that people feel moral responsibility (i.e., universally binding, context-independent, serious, and punishable obligations) toward things that they see as possessing “intrinsic value” but not toward things they see as possessing inherent value, then this distinction would be relevant to the question of what morality is in the descriptive psychological sense we are seeking here. But until we have such proof, I will offer a fuzzy concept of inherence, the meaning of which is intended to precisely match the fuzziness of everyday cognition, as a first approximation. Perhaps the concept can be further nuanced, but for now all I mean to express is the idea that moral experience rests on value that is attributed not to subjective opinion or external conditions but, somehow or someway, to the thing itself—and this, even if philosophical reflection causes us to deny that a “thing in itself” exists—and theorists should perhaps rely on further empirical evidence to support any finer parsing.

Given the above articulation of the proper essence of morality, we can see that morality governs the normative relations between perceived moral agents and all the people, places, creatures, objects, ideas, ways of life, and so on, that are understood as possessing sufficient inherent value (both positive and negative). Morality is not necessarily restricted to dyadic relations between agents and vulnerable patients. It also does not always function to promote cooperation or necessarily involve harm or negative affect. More to the point, we can see precisely how morality precedes moral values. For a moral value to emerge, governing our treatment of inherently valuable things, we must already have attributed inherent value to the things toward which we feel morally responsible. Thus, the real irreducible basic elements of moral cognition are not moral values but evaluative ontological relationships, that is, the various ways in which we frame the things in our world with inherent value attributions—as beautiful, powerful, precious, dangerous, evil, disgusting, or divine; as friends, enemies, family, aliens, predators, rapists, vermin, or heroes; and so on. These ontological framings have immediate moral implications. As our ontological framing changes, so does our sense of moral responsibility—prior to and, to a significant extent, independent of our moral values.

Implications for the Moral Values Debate

I will return to explicating the relation between values and ontological frames in the Conclusion. First, however, I would like to turn back to my critique of the argument between TDM monists and MFT pluralists. How does this definition of the proper essence of morality lead to an intervention in the current debate? Modeling moral cognition and behavior in terms of moral values while ignoring or overgeneralizing the more fundamental ontological relationships from which morality emerges blinds us to important moral phenomena, including interpersonal and intrapersonal dynamics of moral cognition. In this section, I expose blind spots within both monistic and pluralistic theories and show how to address the problem by tracing the dynamics of moral cognition to their ontological roots.

How can we best explain interpersonal and intergroup differences in moral cognition? My claim is that the surest way is not to interrogate people's values but to look instead at their ontologies. If one person sees an entity as being inherently valuable and another views the same entity as a mere instrument with no inherent value, these two individuals will diverge in their views of how the entity ought to be treated—indeed, they will differ in whether or not they feel moral responsibility toward the entity at all. And this divergence need not reflect any difference in their values nor in how moral cognition works for either party. It is simply that each individual stands in a different relation to the entity in question, a different ontological frame. The same logic holds at the group level. As we saw at the beginning of this essay, the fact that a particular culture values justice tells you nothing about which entities it deems worthy of justice or how its view of what is just varies with its specific ontological framings of the subjects of justice. Thus, although I would not deny that individuals and cultures differ in how much they emphasize certain values, I would like to suggest that tracing such differences in people's values should be a secondary form of analysis that complements a more fundamental discussion of ontological framing.

To understand why a framing analysis is important, let us take a closer look at interpretations of moral phenomena made by contemporary monists and pluralists, respectively. I will begin with MFT pluralists. Haidt and colleagues argue persuasively that conservatives emphasize the value of sanctity/purity more than liberals do (Graham, Haidt, & Nosek, 2009; Haidt, 2012). But even if this is true, such information about these groups' values will not always be useful for explaining their behavior. For instance, if it is true that liberals tend to emphasize sanctity less than their conservative

counterparts do, it is also true that there are instances of the inverse, when liberals take up a sanctimonious attitude and conservatives do not. Indeed, Haidt (2012) offers exceptions to this rule, including an anecdote about Western liberals' sanctimonious feelings about eating a balanced breakfast (p. 15). In this case, Haidt suggests that the shoe is on the other foot, with liberals sanctifying an entity that conservatives tend to treat as mundane. But Haidt does not address the implication of his anecdote: In this example, a pluralistic analysis of intergroup differences in values gives us the wrong expectation. If all we know is that conservatives care more about sanctity than liberals, we will expect conservatives to be more sanctimonious about their breakfast than liberals. Thus, to avoid making such an error, we need to dig down below the level of values and look at differences in their ontologies. Now, I am not arguing that Haidt makes this error himself—after all, it is his anecdote. What I am pointing out is that Haidt can avoid the error only if he implicitly acknowledges the relevance of ontological framing: What is decisive for moral judgment is not which abstract values one's culture emphasizes but which specific beings one frames as having inherent value—including all the entities implicated in one's choices about breakfast.

Even when a pluralistic values analysis does not lead to a false prediction, our theoretical interpretation will still be more precise if we ultimately explain moral judgments and behaviors in terms of ontological framing. Let us consider another example that illustrates this more nuanced claim. Haidt et al. (1993) conducted cross-cultural studies that examined how different human groups judge actions that are harmless but that violate certain cultural taboos. One of the stories Haidt et al. told participants was about a person who chooses to cut up his or her own national flag and then to use the pieces as rags to clean a toilet. Haidt et al. showed that certain populations consider this action to be a moral violation, whereas others do not, and they explained this difference in terms of moral values. For those who consider such a use of one's flag immoral, the action seems to violate a value of purity/sanctity, whereas those who consider the action morally innocuous are evidently denying the moral relevance of purity/sanctity and focusing instead on whether there are harmful consequences to the action. Haidt et al.'s interpretation here has evident utility—one may well be able to predict participants' judgments in the flag dilemma on the basis of information about their values. And yet the ability to make predictions is not the only measure of a good theory, and MFT pluralists' interpretation of this finding is nonetheless flawed.

The reason for one population's moralizing interpretation of the flag scenario is not, strictly speaking, that

it values sanctity. After all, one must clean one's toilet with something—and presumably they would not have any moral problem with using regular old rags as long as these rags were not perceived as sanctified objects. Thus, the precise explanation for why this population objects to this use of the flag is not that it values sanctity but rather that it sanctifies this particular object—the flag. Given this ontological framing of the flag as an inherently valuable, sacred entity, even private treatment of the flag will be governed by moral law. Likewise, it is not accurate to say that the population that sees no moral violation in the flag scenario simply does not care about the value of purity/sanctity at all. There are things that members of this population would presumably consider unequivocally wrong for toilet cleaning—but I will leave that to your imagination. Thus, the precise reason why some people do not see the flag scenario as a moral violation is not that they do not value sanctity but that, for them, this flag is not sacred but is merely a means to a valuable end—for instance, symbolic expression, the value of which is also relative to other endeavors that they may or may not view as inherently valuable. Given this ontological framing of the flag as an instrument with no inherent value, one's private treatment of one's private flag has no moral implications. Thus, the correct interpretation of this finding necessarily involves a consideration of ontological framing and should not merely appeal to intercultural differences in values.

The implication of this analysis is that any explanation of moral cognition should consider the relevant ontological framing. And although some cases, such as the flag scenario, reveal rather subtle theoretical problems with pluralists' interpretations of intergroup differences in moral judgment, it is easy to think up cases in which a pluralistic values analysis might lead us more egregiously astray. For instance, if you could ask an early White American one question to discover their moral feelings about slavery, what would it be—a question probing how much they value justice, welfare, or liberty, or a question probing to what extent they view non-Whites as being equal to Whites? It seems obvious to me that the question about values would lead one astray. Just like the Nazis, early supporters of slavery did not deny the value of justice; they simply denied that the enslavement of Africans was unjust. Why? Because of how they ontologically framed people of African ancestry, Whites of European ancestry, the mission of Christianity, the goal of civilization, and so on. This example reveals the importance of a reductive analysis that digs down below the level of moral values to the place from which morality emerges—in our attributions of inherent value (or denial thereof) to things in our world.

Perhaps some readers find this exercise in interpretation convincing but are not sure how this argument might be tested empirically. After all, we cannot go back in time to question Nazis or early White Americans. Fortunately, we do not need to go back in time. Researchers could take any morally contested issue of our day—abortion, illegal immigration, climate change, gun control, transgender rights, and so on—and ask two different kinds of questions: one kind probing participants' values and another kind probing their ontological beliefs. We could then determine which form of questioning was more predictive of participants' moral judgments: framing questions or values questions. For instance, participants could rate on a Likert scale their level of agreement with ontological framing statements such as "Fetuses are persons" or "Climate change is caused by humans" and also with values statements such as "I value personal liberty" or "I value personal responsibility." My prediction is that participants' level of agreement with well-chosen "framing" questions will significantly predict their moral judgments about such issues as abortion and carbon emissions, whereas participants' level of agreement with equally well-chosen values questions will be much less predictive of their moral judgments. However, although I am confident that ontological framing questions would outperform values questions in such a test, I caution that such an experiment would not speak to my most important claim. The goal of moral psychology should not be simply to predict moral judgments but also, more fundamentally, to provide a proper interpretation of the dynamics of moral cognition and its influence on behavior. As I have shown and support further below, such an interpretation necessarily involves tracing moral judgments to their roots in ontological relationships—even if abstract values can sometimes provide indirect information about these ontologies. This is the case because ontological framing is fundamental and necessary to moral cognition, while abstract values are secondary and contingent.

So much for my critique of MFT pluralists' explanations of intergroup variance in moral judgment—what about monism? Let us consider the above scenarios again, this time from the perspective of TDM monists. The liberals' moralizing concerns about a balanced breakfast, our monists might argue, are ultimately all about harm. Are people being exploited by farmers and factory owners involved in a chain of breakfast production? Is the breakfast unhealthy, thereby harming the eater? Is harm being done to animals, plants, or ecosystems, perceived as suffering patients—a phenomenon TDM describes as *dyadic completion*, the idea that we tend to fill in unclear or fuzzy aspects of moral situations to match our dyadic moral template of a

vulnerable patient and an intentional agent (Schein & Gray, 2018)? It seems to me that the TDM explanation works for this case, although I would clarify, contrary to earlier iterations of the theory (e.g., Gray et al., 2012) and in agreement with Schein and Gray (2018), that one need not attribute personhood or mentality to plants and ecosystems to feel morally responsible for them. So far my model is in rough agreement with a TDM model, in the specific case in which the actions of an agent are perceived as harming a patient of some kind or other.

A TDM interpretation is stretched beyond plausibility by Haidt's flag vignette, however. In interpreting this scenario, TDM monists would be forced to assert that people who morally object to cleaning a toilet with one's country's flag are necessarily perceiving the flag itself as a suffering patient or are imagining some other patient who will suffer as a result of this misuse of the flag. Neither of these options seems likely to me, and I have already argued above that such dyadic completion is not strictly necessary for a sense of moral obligation to arise because the precise nature of our moral obligations is determined by the specificities of our ontological relationships, and morality is not restricted to relationships that fit the TDM prototype. However, dyadic completion is distantly possible in this case. Thus, we may ask: In cases in which one's sense of moral responsibility arises toward real or imagined "vulnerable patients," is the TDM explanation more adequate than mine? The answer is that vulnerable patienthood *is* an ontological frame—so in such cases our explanations are basically the same. The advantage of my analysis, however, is that it acknowledges greater variety in the beings that people may value inherently, accommodating not only our moral obligations not to harm but also obligations to obey, respect, believe, or protect, to name just a few. My model thus allows me to explain the flag vignette without making the implausible claim that everyone who considers such a use of the flag to be immoral is necessarily imagining a patient suffering as a result. Although some people might perform dyadic completion in this case, it seems likely that most of those who judge the flag scenario to be morally wrong are not imagining harm being done to any patient—they simply feel it is wrong to treat a sacred object without the proper reverence. My model has no problem accommodating all cases, responding to the unique moral implications of diverse ontological relationships. It is true that we have moral relationships with vulnerable patients, but there are other kinds of moral relationships too.

In addition to the anecdotal case above, I believe there is already research into interindividual and intergroup moral variance that supports my interpretation while

causing problems for any model that relies on abstract values as the fundamental unit of explanation. Here I would draw particular attention to Shweder and colleagues' important work on intercultural variation in moral judgment (e.g., Shweder, 1990, 2015, 2017; Shweder et al., 1987, 1997). Although Shweder's theory of the Big Three ethical domains of autonomy, community, and divinity greatly influenced MFT pluralism, Shweder is not committed to describing moral differences in terms of abstract values, and his studies illustrate precisely the relationship between ontological framing, moral judgments, and moral values. Shweder's review of an earlier draft of this essay is worth quoting at length, as it shows the way to synthesizing my own model with a model of value pluralism. In this quote, Shweder is interpreting his own seminal study (Shweder et al., 1987), in which he and colleagues interviewed participants in the Hindu temple town of Bhubaneswar, Orissa, India, as well as others in Hyde Park, Illinois, asking Turiel-style questions about the appropriateness of various everyday activities. Shweder et al. famously found that members of these different populations had strikingly different moral intuitions about a wide range of behaviors, showing that some non-Western populations moralize a much broader set of objects and situations than psychologists had been willing to recognize. Describing to me the stark differences in Indian and American participants' responses to a question about whether it is okay for a widow to eat fish, Shweder points out how these responses suggest an interrelatedness of ontological framing and moral values, where the framing comes first and then gives rise to value attachments that are specific to one's status and station within a given tradition (see also D'Andrade, 2008):

The ontological framing certainly produces dramatic differences in moral judgment but the content of the relevant frames (the positing of a status obligation associated with the station of widowhood and the continued mutual reliance of husband and wife even after death versus the obligation to protect the personal liberties and areas of discretionary choice of autonomous individuals) results in a personal attachment (associated with that status or station in life in this particular tradition) to values (for example, loyalty versus liberty) that seem very much like a part of a package deal for moral cognition (and motivation). (R. Shweder, personal communication, June 24, 2019)

I heartily endorse this outline of the dynamics of moral cognition: Ontological framing produces moral judgments that can subsequently give rise to interpersonal and intercultural differences in our attachments

to certain moral values, the meaning of which take shape within particular traditions. Such an interpretation shows the harmony between my model of ontological framing and Shweder's pluralism. I would only want to nuance the suggestion that ontological frames and moral values are "part of a package deal" by emphasizing the hierarchical organization of this package: One cannot develop moral values until one has already begun to frame things as being inherently valuable, and one can dramatically alter one's moral sensibility through ontological reframing even as one's values remain the same (recall Berlin, 1998, and see also my discussion of intraindividual moral dynamics below). The intercultural differences in abstract values noted above were neither fundamental nor categorical. Shweder's Indian respondents were not wholly indifferent to liberty, nor were the American respondents without any sense of loyalty. Rather, this example shows that fundamental differences in ontological framing give rise to disagreements over *what counts* as loyalty or disloyalty, liberty, or oppression.

So much for my analysis of moral differences between individuals and groups. What about the dynamics of moral cognition within individuals? Values theories have little to say about such dynamics—at least on rapid time scales—simply because our abstract values do not tend to change very rapidly. Indeed, people see their core moral values as their most stable source of identity, feeling that if these values were to change they would no longer be the same person (Heiphetz, Strohminger, & Young, 2016; Strohminger & Nichols, 2014, 2015). Ontological framing can vary much more rapidly than values, however. Although we maintain some ontological judgments pretty rigidly, other ontological frames warp with transient emotions and moods. Thus, if an ontological analysis has an advantage for predicting and interpreting interpersonal and intergroup differences in moral cognition and behavior, it has an even greater advantage at the intrapersonal level.

Consider the following anecdote. In a fit of rage, one may call their friend bad names, yell, or even grow violent. In the moment, we may feel justified in doing things that we would never consider while happy or peaceful. Our values probably do not change when we get angry, so how can our sense of what is morally proper shift so dramatically? My model offers a simple answer: Our ontological framings of other people, places, creatures, objects, ideas, and ways of life can vary with shifting attention, memory, understanding, emotion, or mood, and such ontological shifts have immediate moral implications. For instance, name calling and other forms of derogation are explicit ontological reframes—e.g., "you are such a piece of shit"—that can transform our sense of morally proper behavior,

even as our values remain unquestioned and unchanged. Political pundits talk about how derogatory language creates a *permission structure* for behaviors that would otherwise seem immoral. I agree. Whether such language is deployed strategically by politicians and propagandists or arises silently within our own minds during a moment of rage, this permission structure is nothing but an ontological reframing that alters our sense of right and wrong—sometimes transiently, sometimes in more lasting ways. A flash of anger that bends one's ontological frames is just one striking example of the way that moral cognition always works. Praise is also a mode of ontological framing. A rush of warm feelings or empathy can change how we view our opponents or enemies in a moment: With softened hearts, we may no longer feel they deserve our aggression. Such emotionally motivated reframings may be transient. But our ontological frames can also solidify into lasting prejudices or ideologies. The labels we use to identify ourselves or other races and genders; to speak about nonhuman animals, our planet, our universe, our sacred ideals; or to specify our relationships to all of these things and more shape our sense of morally proper behavior at every moment, prior to and independent of the emergence of moral values. And although TDM monists and MFT pluralists both emphasize that affects and emotions influence moral cognition, they do not specify the basic way in which this effect is mediated. My model offers a specific answer—*affects and emotions affect moral judgment, insofar as they do so, by bending our ontological frames.*

Describing moral cognition in terms of ontological relations that are subject to both stability and rapid change gives my model a flexibility that values theories lack. Only such a flexible model is equal to the task of explaining the dynamics of everyday moral cognition within individual minds. Is this claim testable? I do not see why the power of an ontological analysis to account for intraindividual moral fluctuations should not be empirically comparable to that of values theories. My prediction is that values theories will be largely blind to intraindividual moral fluctuations because people's values do not change so quickly. However, well-designed studies should reveal a strong correlation between changes in our sense of moral responsibility and fluctuations in ontological framing because moral cognition is necessarily mediated by ontological framing.

In fact, I think that many studies already exist that strongly support my ontological interpretation of intraindividual moral dynamics while resisting any interpretation in terms of values, whether one, several, or many. Consider, for example, Milgram's famous experiments from the 1960s in which he showed that everyday people are willing to shock innocent strangers at what they

believe to be excruciatingly painful and dangerous voltages when instructed to do so by an authority figure (Milgram, 1963). In the classic interpretation, Milgram's findings are a powerful illustration of the all-too-human willingness to obey authority, even when such obedience goes against our own moral convictions. However, a closer look at the evidence reveals that the classic interpretation of this finding is flawed—study participants were not blindly or passively obedient to authority, as often assumed. For instance, both in Milgram's original experiments and in a more recent replication (Burger, 2009), participants were universally resistant to the prompt experimenters sometimes used as a last attempt to induce cooperation: "You have no other choice. You must go on" (Reicher & Haslam, 2011, pp. 167–168). No one ever obeyed this command but chose instead to rebel in the face of explicit coercion. If not blind obedience to external authority, what induced participants to engage in such shocking behavior? Reicher and Haslam (2011) argue that, rather than being passively bent by the wind of authority, participants responded to numerous cues that shifted their alliances in the context of the experiment. For instance, compliance with the direction of the experimenter was reduced by all factors that took subjects out of the ontological frame of shared participation in an important scientific project, whether by emphasizing their subjection to the authority of the experimenter or by, in various ways, delegitimizing and devaluing the experiment. Factors that helped place participants in a common frame with the shock victim, such as increasing their physical proximity, also decreased compliance. Conversely, compliance was increased by factors that emphasized subjects' participation, together with the experimenter, in an inherently valuable scientific project.

Thus, it would seem that Milgram achieved his most dramatic effects through at least three forms of deliberate reframing: attenuating the sense of proximity between participants and the person they believed they were shocking, strengthening participants' identification with an inherently valuable scientific project, and buttressing participants' sense of the inherent value of the research they were supporting. These behavioral effects were thus mediated by participants' internal moral capacity, as they responded to numerous experimental manipulations of value and of their own connection to valuable things. Thus, in the situation in which they found themselves, most participants were not simply abnegating their conscience out of a lack of character strength or personal willpower. Rather, they found themselves in a situation in which their obligations as participants in an important (inherently valuable) scientific experiment conflicted with their obligations to behave compassionately toward another

(inherently valuable) study participant. Compliance with the direction to shock the other participant was not primarily an abnegation of moral responsibility but a dark moral choice, a performance of moral responsibility in a morally complex situation. And this choice—insofar as it was a moral choice and not a simple abnegation of responsibility in the face of authority—appears to have been mediated by ontological framings that shaped participants' sense of morally proper behavior.

Although a plurality of conflicting values may have informed participants' choices in this experiment—we might note a conflict between the moral foundations of care for others and respect for authority, for instance—we should note that Milgram's findings cannot be explained in terms of this conflict, as there is no good reason to believe that his various experimental conditions changed subjects' values. Thus, the much more plausible interpretation is that Milgram's experimental conditions each framed differently the relation between the participant and the other things toward which participants felt a sense of moral obligation, and these experimental manipulations of ontological framing altered participants' sense of right and wrong.

Second, consider a study that Haidt himself leans on in his opus *The Righteous Mind* (Haight, 2012). Sherif, Harvey, White, Hood, and Sherif (1954/1961) studied the effects of group identification on behavior in groups of young boys in the 1950s. By placing individuals in competing groups, Sherif et al. found that they could activate pro-in-group and anti-out-group behaviors in their young participants. As Haidt (2012) describes it,

Both sides [spontaneously] created flags and hung them in contested territory. They destroyed each other's flags, raided and vandalized each other's bunks, called each other nasty names, made weapons (socks filled with rocks), and would often have come to blows had the counselors not intervened. (p. 175)

Haidt uses this experiment as a compelling example of human "groupishness": our capacity to work together as a group while suppressing individual selfishness, often in the context of intergroup competition. However, I would like to point out another implication of this study: Although the observed behavioral changes cannot be explained in terms of values, they can be easily explained in terms of ontological framing. Considering that the experiment was fairly brief and involved no program of moral education, it would be highly implausible to suppose that the observed behavioral effects were induced by a sudden change in the

boys' values. If, then, we reject this hypothesis, we can see what is right in front of us: What changed is the ontological framing. Whereas before these boys might have encountered each other as individual strangers on the street, once the experiment began they became members of competing groups and began to see each other as enemies. And the proper way to treat an enemy is very different from the proper way to treat a random person on the street. Thus, as with Milgram's experiments, the effect in Sherif et al.'s experiment was plausibly mediated not by modification of the boys' values but by a shift in their ontological framing of one another.

These are just two prominent examples out of many that could be adduced in support of my model. For instance, the above interpretation of Milgram's experiments can be made of most experiments whose conditions modulate moral judgments and behaviors within individual participants, especially when such experiments are brief enough to render any change in abstract values implausible. It seems to me that there are generally two possible explanations for any valid finding that meets these criteria: Either the experiment suggests an absence of moral engagement where we would generally expect moral cognition to be operating (as in the classic interpretation of Milgram), or the experiment points to the modulation of moral cognition by experimental manipulations of ontological framing. By contrast, only in longitudinal studies, such as those of Kohlberg (1984) and Gilligan (1982), do we observe meaningful intraindividual alterations of participants' moral values.

In addition to the classic situationist studies discussed above, recent research into dehumanization (Harris & Fiske, 2006, 2011; Haslam, 2006; Haslam & Loughnan, 2014; Leyens et al., 2003; Leyens, Demoulin, Vaes, Gaunt, & Paladino, 2007) and superhumanization (Waytz, Hoffman, & Trawalter, 2015; Wilson, Hugenberg, & Rule, 2017) also powerfully illustrates the moral relevance of both implicit and explicit forms of ontological framing while resisting any account that relies on abstract values. To illustrate the gist of this work, we may simply reconsider the epigraph of this essay, in which Berlin grounds the Nazis' moral ideology and behavior in their beliefs about "the facts." Many of these beliefs are paradigm cases of dehumanization, as the Nazis generated moral support for their cause by framing people of color, people with disabilities, nonheterosexuals, Jewish people, and other populations as being, in various ways, dangerous and degenerate. More recently, theorists have also pointed to the unsettling moral implications of more subtle, everyday forms of dehumanization, or *infrahumanization* (e.g., Leyens et al., 2003; Leyens et al., 2007), and to problems that emerge when we superhumanize Black men, attributing

to them superhuman strength and imagining that they feel less pain than others (e.g., Waytz et al., 2015; Wilson et al., 2017). Such ontological framings illustrate the dark side of morality, as our beliefs about "the facts" promote discrimination, prejudice, hatred, fear, war, genocide, slavery, and the like. And crucially, these framings can change one's sense of right and wrong without affecting one's abstract values. The belief that an individual or group is dangerous, demonic, or degenerate has immediate moral implications, regardless of what values one holds.

TDM monists have had success explaining dehumanization as a way of attenuating the dyadic moral relationship. To a certain extent, I agree with their analysis and would just add two caveats. First, as I have emphasized above, the moral dyad between an agent and vulnerable patient is a particular case within my more general model. Second, TDM monism reduces morality to the abstract principle of doing no harm to moral patients. However, dehumanization may also diminish our sense that we ought to show respect or reverence, or it may change our idea of what is fair and just, or shift our loyalties, and so on. This complexity of the implications of dehumanization emerges not from the fact that there are a variety of ways to harm vulnerable patients but rather from the fact that we view others through diverse ontological frames. Humans are more than vulnerable patients, and so dehumanization does more than attenuate our responsibility to refrain from harming the vulnerable.

So much for the evidence. Both classic situationist findings and recent studies of dehumanization strongly support the claim that we ought to explain interindividual, intergroup, and intraindividual dynamics of moral cognition and behavior in terms of ontological framing, and this evidence also reveals insufficiencies and blind spots in theories that explain morality in terms of abstract values. Ontological framing is ubiquitous—wherever there is moral cognition, it is grounded in ontological interpretations of what is inherently valuable and what is not. Thus, we might also use ontological framing to understand the moral implications of sexual or medical objectifications, historical narratives and counter-narratives, the beliefs underlying political ideologies, hero worship and celebrity, and the claims and stories through which religions and cults govern followers. Of course, my model calls for further empirical investigation. However, even if it should somehow be shown that, in some cases, moral values are better predictors of moral judgments and behaviors than are ontological frames, the essence of my critique would remain unassailed: Morality is always necessarily determined by ontological frames—the real irreducible basic elements of the moral domain—prior to and, to a significant degree, independent of the action of moral values.

Conclusion

Let us take one last look at the details of the model I am proposing, with special attention to the meaning of my central concept. Even readers who are sympathetic to my argument may wonder why have I chosen the term *ontological framing*. For instance, one may ask why, given the importance of the distinction between attributed value and principles, I have not framed this more simply in terms of a relation between ontological attributes and moral principles. There are a few reasons. One is that in everyday thought and language, we do not always break entities down into piles of attributes. Instead, we use holistic language: ontological labels that vaguely define something or someone, often through metaphor or analogy. And these labels carry with them evaluative resonances with immediate and wide-ranging moral implications. We can observe the holistic nature of ontological framing, for instance, in dehumanizing slurs. If I describe a group of people by calling them “vermin,” listeners probably do not immediately break down the ontological frame *vermin* into a bunch of smaller attributes. Instead, if you are persuaded by my hateful speech, you probably contract some vague feeling that it is appropriate for “us” to treat “them” like the “vermin” they are. Of course, I am cognizant that we do think in terms of attributes sometimes, and I do not deny the relevance of such thinking for moral psychology. But this is not a problem for my concept. Breaking down an entity into attributes is also a way of ontologically defining the entity—so it is just another kind of ontological frame.

Another reason I did not want to describe the value carried by ontological frames as an attribute is that this would fail to convey the relational character of ontological understanding. In everyday understanding, the “being” of an entity often expresses a relationship between things while smuggling in implicit evaluations of the relationship and implicit obligations for how we ought to behave to fulfill our particular roles. For instance, if I call you “friend,” you probably hear positive resonances in this ontological framing of our relationship while perhaps feeling yourself obligated to perform certain duties and expecting to enjoy the rights accorded to you as my friend. If you badly betray your duties as a friend, I may start to consider you my enemy. I do not need to call you a *bad* enemy because the frame *enemy* already carries negative resonances, and I implicitly understand that my obligations toward you as my enemy are different from those I felt toward you when I considered you a friend. Your ontological status as my friend or enemy is not reducible to the value I attribute to you—although these frames carry evaluative resonances—nor is being a friend or an enemy an

attribute that you possess. Rather, these ontological frames establish the shape of the relationship between us, reciprocally defining us, while simultaneously smuggling in implicit expectations for how we ought to treat each other. We know our roles, so to speak: If you are my friend, I am yours; and if you become my enemy, I more than likely transform into your enemy. And we will likely feel different obligations toward each other from within these distinct ontological frames, regardless of what abstract values we espouse. Thus, one could say that ontological frames hold both the windows through which we view others and the mirrors within which we perceive ourselves. The ontological frame *friend* contains both window and mirror.

Finally, I chose the term ontological framing because it allows morality to be as complex and dynamic as it actually is. One’s attributes do not change rapidly and constantly, and neither do one’s values. However, we all inhabit many ontological frames, and those that define our relationships to other entities may change in an instant. For instance, at the moment I am a brother, son, researcher, citizen, boyfriend, and so on. These labels are not attributes that I possess but frames that specify my relationship to my siblings, my parents, the research community, my nation of birth, my girlfriend, and so on, while also binding me with moral obligations. Breaking my obligations can therefore dissolve the frame, changing my identity, my role. A bad son or brother might be disowned; an unethical researcher might be fired or blacklisted; bad citizens could have their citizenships revoked; a bad boyfriend could be dumped. Each of these actions would involve the destruction of some ontological frame, the changing of a relationship that defines who I am in relation to other things and also the remaking of the rights and responsibilities that had been built into those frames. Once you have dumped me, it may no longer be appropriate for me to greet you with a passionate kiss—I am no longer your boyfriend, with all the rights and responsibilities built into that frame; once I have been fired I may no longer be allowed to come in to work—I am no longer an employee.

Frames can change immediately in this way, through performative speech or action, such as a breakup or a firing. But frames also change at an implicit level, as they fluctuate with the dynamics of our internal psyches. As already described, my sense of who you are can bend with transient emotions and moods. In a moment of anger, I might see you momentarily as my enemy or as a piece of shit. I can also forget you, and when I do, you fall through all my ontological frames and thereby escape my moral consideration. Thus, because we always inhabit many ontological frames, and these frames are subject to many internal and external sources

of immediate change, our sense of morally appropriate behavior is incredibly dynamic. Moral pluralists may attempt to explain our moral complexity by appealing to the fact that people have multiple irreducibly distinct values. But we should appreciate that our values are fairly static, and the complexity that emerges from conflicts in values is vanishing compared with the moral complexity that we negotiate every day simply by being who and what we are.

This, in brief, is my argument. I do support the move away from monistic models, in which morality is supposed to be founded on a single principle, and toward a pluralistic model acknowledging the irreducible complexity of the moral domain. But I think MFT pluralists' metaphor of values as foundations supporting complex moral cities needs to be turned upside down. In fact, moral relationships emerge organically before we ever develop moral values. Only on the surface of all of this dynamism and complexity, in communication with other individuals and groups, do we develop abstract values that help us navigate diverse relationships, situations, and roles. Such values are certainly important, but they should not be described as moral foundations or treated as irreducible basic elements of the moral domain.

Action Editor

Laura A. King served as action editor for this article.

ORCID iD

Bree Beal  <https://orcid.org/0000-0001-6544-302X>

Acknowledgments

I thank Elspeth Brotherton, Guga Gogia, Lily Oster, Philippe Rochat, Richard Shweder, Mike Winters, and others who influenced this manuscript in various ways.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Notes

1. Monists accept pluralists' account of morality's function over the course of evolutionary time (Schein & Gray, 2018, p. 14).
2. Pluralists also accept the claim that interpersonal harm is central to moral cognition (Haidt et al., 2015).
3. We might also nuance this third condition to accommodate Turiel's criteria of seriousness and punishability. These criteria might be described as subjectively quantitative. They suggest that the inherent value we attribute to something, as well as the badness of our treatment, must surpass a certain threshold before we consider mistreatment of that thing to be a moral issue. This threshold is necessarily subjective. I might consider

an ant to have inherent value, but I would not consider the killing of an ant to be immoral because I do not value the ant so highly; however, some might consider the ant's life sufficiently valuable that to kill it would be immoral. In this essay, everything I say about inherent value presumes that it surpasses one's subjective morality threshold.

References

- Ardila-Rey, A., & Killen, M. (2001). Middle-class Colombian children's evaluations of personal, moral, and social conventional interactions in the classroom. *International Journal of Behavioral Development*, 25, 246–255.
- Berlin, I. (2001). My intellectual path. In H. Hardy (Ed.), *The power of ideas* (pp. 1–23). Princeton, NJ: Princeton University Press.
- Bersoff, D. M., & Miller, J. G. (1993). Culture, context, and the development of moral accountability in judgments. *Developmental Psychology*, 29, 664–676.
- Burger, J. (2009). Replicating Milgram: Would people still obey today? *American Psychologist*, 64, 1–11.
- Carey, N., & Ford, M. (1983, August). *Domains of social and self-regulation: An Indonesian study*. Paper presented at the meeting of the American Psychological Association, Los Angeles, CA.
- D'Andrade, R. (2008). *A study of personal and cultural values: American, Japanese, and Vietnamese*. New York, NY: Palgrave Macmillan.
- Darwin, C. (1871). *The descent of man and selection in relation to sex*. London, England: John Murray.
- Essence. (n.d.) In *Merriam-Webster's online dictionary*. Retrieved from <https://www.merriam-webster.com/dictionary/essence>
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development* (2nd ed.). Cambridge, MA: Harvard University Press.
- Graham, J. (2015). Explaining away differences in moral judgment: Comment on Gray and Keeney (2015). *Social Psychological & Personality Science*, 6, 869–873. doi:10.1177/1948550615592242
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 47, pp. 55–130). San Diego, CA: Academic Press. doi:10.1016/B978-0-12-407236-7.00002-4
- Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitsch, C., & Mooijman, M. (2018). Moral foundations theory: On the advantages of moral pluralism over moral monism. In K. Gray & J. Graham (Eds.), *The atlas of moral psychology: Mapping good and evil in the mind* (pp. 211–222). New York, NY: Guilford Press.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96, 1029–1046.
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101, 366–385.
- Gray, K. (2014). Harm concerns predict moral judgments of suicide: Comment on Rottman, Kelemen and Young

- (2014). *Cognition*, 133, 329–331. doi:10.1016/j.cognition.2014.06.007
- Gray, K., & Keeney, J. E. (2015a). Disconfirming moral foundations theory on its own terms: Reply to Graham (2015). *Social Psychological and Personality Science*, 6, 874–877.
- Gray, K., & Keeney, J. E. (2015b). Impure or just weird? Scenario sampling bias raises questions about the foundations of morality. *Social Psychological and Personality Science*, 6, 859–868.
- Gray, K., Schein, C., & Ward, A. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, 143, 1600–1615. doi:10.1037/a0036149
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23, 101–124.
- Greene, J. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. New York, NY: Penguin Books.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon Books.
- Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about morality. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 2: The cognitive science of morality* (pp. 181–217). Cambridge, MA: MIT Press.
- Haidt, J., Graham, J., & Ditto, P. (2015, October 28). The Volkswagen of moral psychology [Blog post]. Retrieved from <http://www.spsp.org/blog/dyadic-morality>
- Haidt, J., & Kesebir, S. (2010). Morality. In S. Fiske & D. Gilbert (Eds.), *Handbook of social psychology* (5th ed., pp. 797–832). Hoboken, NJ: John Wiley.
- Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 613–628.
- Harris, L., & Fiske, S. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, 17, 847–853.
- Harris, L., & Fiske, S. (2011). Dehumanized perception: A psychological means to facilitate atrocities, torture, and genocide? *Zeitschrift fur Psychologie*, 219, 175–181.
- Haslam, N. (2006). Dehumanization: An integrated review. *Personality and Social Psychology Review*, 10, 252–264.
- Haslam, N., & Loughnan, S. (2014). Dehumanization and inhumanization. *Annual Review of Psychology*, 65, 399–423. doi:10.1146/annurevpsych-010213-115045
- Heidegger, M. (2010). *Being and time* (J. Stambaugh & D. Schmidt, Trans.). Albany: State University of New York Press. (Original work published 1927)
- Heiphetz, L., Strohminger, N., & Young, L. (2016). The role of moral beliefs, memories, and preferences in representations of identity. *Cognitive Science*, 41, 744–767.
- Henrich, J., & Gil-White, F. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution & Human Behavior*, 22, 165–196. doi:10.1016/S1090-5138(00)00071-4
- Hollos, M., Leis, P., & Turiel, E. (1986). Social reasoning in Ijo children and adolescents in Nigerian communities. *Journal of Cross-Cultural Psychology*, 17, 352–376.
- Kant, I. (2002). *Groundwork for the metaphysics of morals* (A. Wood, Ed. & Trans.). New Haven, CT: Yale University Press. (Original work published 1785)
- Kelly, D., Stich, S., Haley, K. J., Eng, S. J., & Fessler, D. M. T. (2007). Harm, affect, and the moral/conventional distinction. *Mind & Language*, 22, 117–131. doi:10.1111/j.1468-0017.2007.00302.x
- Kohlberg, L. (1971). Indoctrination versus relativity in value education. *Zygon*, 6, 285–310.
- Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of justice* (Vol. 1). New York, NY: Harper & Row.
- Kohlberg, L. (1984). *The psychology of moral development: The nature and validity of moral stages* (Vol. 2). New York, NY: Harper & Row.
- Korsgaard, C. (1983). Two distinctions in goodness. *Philosophical Review*, 92, 169–195.
- Lewis, C. (1946). *An analysis of knowledge and valuation*. La Salle, IL: Open Court.
- Leyens, J.-P., Cortes, B., Demoulin, S., Dovidio, J., Fiske, S., & Gaunt, R. (2003). Emotional prejudice, essentialism, and nationalism: The 2002 Tajfel lecture. *European Journal of Social Psychology*, 33, 703–717.
- Leyens, J.-P., Demoulin, S., Vaes, J., Gaunt, R., & Paladino, M. (2007). Infra-humanization: The wall of group differences. *Journal of Social Issues Policy Review*, 1, 139–172.
- Madden, T. (1992). *Cultural factors and assumptions in social reasoning in India* (Unpublished doctoral dissertation). University of California, Berkeley, Berkeley.
- May, J. (2018). *Regard for reason in the moral mind*. New York, NY: Oxford University Press.
- Mayr, E. (1961). Cause and effect in biology. *Science*, 134, 1501–1506.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67, 371–378.
- Miller, J. G., & Bersoff, D. M. (1992). Culture and moral judgment: How are conflicts between justice and interpersonal responsibilities resolved? *Journal of Personality and Social Psychology*, 62, 541–554.
- Moore, G. E. (1903). *Principia ethica*. Cambridge, England: Cambridge University Press.
- Nucci, L. (2001). *Education in the moral domain*. Cambridge, England: Cambridge University Press.
- Nucci, L., Camino, C., & Milnitsky-Sapiro, C. (1996). Social class effects on Northeastern Brazilian children's conceptions of areas of personal choice and social regulation. *Child Development*, 67, 1223–1242.
- Nucci, L., Turiel, E., & Encarnacion-Gawrych, G. (1983). Children's social interactions and social concepts in the Virgin Islands. *Journal of Cross-Cultural Psychology*, 14, 469–487.
- Patil, I., & Silani, G. (2014). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, 5, Article 501. doi:10.3389/fpsyg.2014.00501
- Piaget, J. (1965). *The moral judgment of the child* (M. Gabain, Trans.). New York, NY: The Free Press. (Original work published 1948)

- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Belknap Press.
- Reicher, S., & Haslam, A. (2011). After shock? Towards a social identity explanation of the Milgram "obedience" studies. *British Journal of Social Psychology*, 50, 163–169.
- Rottman, J., Keleman, D., & Young, L. (2014). Purity matters more than harm in moral judgments of suicide: Response to Gray (2014). *Cognition*, 133, 332–334. doi:10.1016/j.cognition.2014.06.008
- Royzman, E., Kim, K., & Leeman, R. (2015). The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. *Judgment and Decision Making*, 10, 296–313.
- Schein, C., & Gray, K. (2015). The unifying moral dyad: Liberals and conservatives share the same harm-based moral template. *Personality and Social Psychology Bulletin*, 41, 1147–1163. doi:10.1177/0146167215591501
- Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, 22, 32–70. doi:10.1177/10888683176982
- Sherif, M., Harvey, O. J., White, B. J., Hood, W., & Sherif, C. (1961). *Intergroup conflict and cooperation: The robbers cave experiment*. Norman: University of Oklahoma Institute of Group Relations. (Original work published 1954)
- Shweder, R. (1990). In defense of moral realism: Reply to Gabennesch. *Child Development*, 61, 2060–2067.
- Shweder, R. (1991). *Thinking through cultures*. Cambridge, MA: Harvard University Press.
- Shweder, R. (2015). Foreword. In L. Jensen (Ed.), *Moral development in a global world* (pp. xii–xviii). Cambridge, England: Cambridge University Press.
- Shweder, R. (2017). Social intelligence in a multicultural world: What is it? Who needs it? How does it develop? In N. Budwig, E. Turiel, & P. D. Zelazo (Eds.), *New perspectives on human development* (pp. 313–329). Cambridge, England: Cambridge University Press.
- Shweder, R., Mahapatra, M., & Miller, J. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1–83). Chicago, IL: University of Chicago Press.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, and divinity), and the "big three" explanations of suffering. In A. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119–169). New York, NY: Routledge.
- Smetana, J. G. (2006). Social domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–153). Mahwah, NJ: Erlbaum.
- Smetana, J. G., Jambon, M., & Ball, C. (2014). The social domain approach to children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 23–45). New York, NY: Psychology Press.
- Smith, A. (1904). *An inquiry into the nature and causes of the wealth of nations* (E. Cannan, Ed.) (2 Vols.). London, England: Methuen. (Original work published 1776)
- Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.
- Song, M., Smetana, J., & Kim, S. (1987). Korean children's conceptions of moral and conventional transgressions. *Developmental Psychology*, 23, 577–582.
- Strohming, N., & Nichols, S. (2014). The essential moral self. *Cognition*, 131, 159–171.
- Strohming, N., & Nichols, S. (2015). Neurodegeneration and identity. *Psychological Science*, 26, 1469–1479.
- Tinbergen, N. (1963). On the aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20, 410–433.
- Tisak, M. S. (1995). Domains of social reasoning and beyond. In R. Vista (Ed.), *Annals of child development* (Vol. 11, pp. 95–130). London, England: Jessica Kingsley.
- Trivers, R. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46, 35–57. doi:10.1086/406755
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, England: Cambridge University Press.
- Turiel, E. (1998). The development of morality. In W. Damon & N. Eisenberg (Eds.), *The handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 863–932). New York, NY: Wiley.
- Turiel, E. (2002). *The culture of morality: Social development, context, and conflict*. Cambridge, England: Cambridge University Press.
- Turiel, E. (2006). Thought, emotions, and social interactional processes in moral development. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 7–35). Mahwah, NJ: Erlbaum.
- Turiel, E., Killen, M., & Helwig, C. (1987). Morality: Its structure, function, and vagaries. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 155–243). Chicago, IL: University of Chicago Press.
- Waytz, A., Hoffman, K. M., & Trawalter, S. (2015). A superhumanization bias in Whites' perceptions of Blacks. *Social Psychological & Personality Science*, 6, 352–359.
- Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology*, 113, 59–80.
- Wittgenstein, L. (1997). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford, England: Blackwell. (Original work published 1953)
- Yau, J., & Smetana, J. G. (2003). Conceptions of moral, social-conventional, and personal events among Chinese preschoolers in Hong Kong. *Child Development*, 74, 647–658.
- Zimba, R. F. (1987). *A study of forms of social knowledge in Zambia* (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN.