



The nonmoral conditions of moral cognition

Bree Beal

To cite this article: Bree Beal (2021) The nonmoral conditions of moral cognition, Philosophical Psychology, 34:8, 1097-1124, DOI: [10.1080/09515089.2021.1942811](https://doi.org/10.1080/09515089.2021.1942811)

To link to this article: <https://doi.org/10.1080/09515089.2021.1942811>



Published online: 21 Jun 2021.



Submit your article to this journal [↗](#)



Article views: 112



View related articles [↗](#)



View Crossmark data [↗](#)



The nonmoral conditions of moral cognition

Bree Beal

Department of English, Clemson University, Clemson, SC, USA

ABSTRACT

Theorists seeking evidence of moral cognition – whether in human infants, nonhuman animals, or any other population – would benefit from a minimalistic description of what moral cognition is. However, such a definition has proven elusive. Some argue that debates over the existence (or not) of moral cognition in various populations turn on unresolvable semantic disagreement over how to characterize the moral domain. I acknowledge a semantic dimension to some disputes and identify another problem: Often, while sidestepping semantics, researchers rely on logically circular operationalizations, defining moral cognition in terms of elements that are already implicitly understood to be features of moral cognition, while failing to answer the question of what makes these features, or their combination, uniquely moral. The present article proposes a single solution to both problems. The issue of semantics is addressed by the identification of a naturally emerging and distinctive cognitive modality that is necessary to all definitions of moral cognition. The problem of circularity is overcome by a reduction of moral cognition to elements that are, in themselves, nonmoral. I call this distinctive combination of nonmoral elements the “molecular” structure of moral cognition.

ARTICLE HISTORY

Received 22 June 2019

Accepted 9 June 2021

KEYWORDS

Early development of morality; moral cognition; morality in nonhumans; moral judgment; moral molecule; moral psychology

Our account of this science will be adequate if it achieves such clarity as the subject-matter allows; for the same degree of precision is not to be expected in all discussions, any more than in all products of handicraft.

–Aristotle, *Ethics*

1. Introduction: What counts as moral cognition?

Suppose you live in an ancient culture where the brightest minds believe that the universe is made of four elements – earth, water, air, and fire. Seeking to define the element “water” for the purpose of scientific investigation, you might start by naming the most obvious properties that distinguish this element from the other three – perhaps beginning with the observation

that water is a liquid. This definition has obvious limitations, however. If you define water as a liquid, you have no criterial basis for distinguishing water from nonwater liquids, and your definition also excludes nonliquid forms of water, such as ice and water vapor. Accordingly, you might refine your definition with additional properties. To distinguish water from juice and honey, you might note that water is comparatively clear, odorless, and tasteless. And you might also observe that water takes the form of a gas, a liquid, or a solid at different temperatures, while remaining essentially itself. These additional properties would lend nuance to your definition, yet you would still be unable to distinguish water from other clear, odorless, tasteless liquids; and as you continued to add properties in an ad hoc manner, your operationalization of water would grow unwieldy.

From our vantage in the present day, it is evident that these difficulties with defining the “element” water are not endemic to the phenomenon itself, but instead reflect the immaturity of scientific understanding at a certain historical moment. Although water does take diverse forms and does mix with diverse substances, it does not follow that it must be defined through an unwieldy list of properties. Key to modern understanding is the appreciation that water is an emergent phenomenon that is built of more basic elements. Although the discovery of the molecule H_2O would eventually bring about new sources of definitional ambiguity, stemming from the further observation that there are diverse, closely related chemical compounds within a typical sample of water (see Weisberg, 2005), these new difficulties are of an entirely different order than those faced by the ancients. It was a major scientific advance when we finally understood that water has a molecular structure entailing simpler elements – elements that are, in themselves, not water.

It seems to me that the young science of moral psychology is in a similar predicament to that of ancient researchers investigating the nature of water – and also that there is an analogous solution that has yet to be discovered. We are currently engaged in building, in a more or less ad hoc manner, taxonomies of different kinds of moral judgments, moral rules, moral values, and so on. And we currently define moral cognition in terms of various salient properties that we associate with morality. But we have yet to identify precisely what moral cognition is made of.

The present article seeks to address this problem by identifying the most basic form of moral cognition and showing how this “molecular” structure derives from more primordial “atomic” elements of cognition. This analogy between physical and cognitive elements is, of course, not perfect. The most crucial difference is that perceptions of morality are fluid and subjectively variable – a problem I discuss in detail (see Section 3). I ask readers to bear in mind my epigraph, taken from Aristotle’s *Nicomachean Ethics*: We should only aspire to the level of precision of which a subject admits. We

may never be able to identify moral cognition as robustly as we can identify a water molecule. Nevertheless, it is possible to approach our subject with more precision than we do currently.

Although the comparison is imperfect, the proposed elements of moral cognition are, in multiple senses, analogous to the physical atoms that make up a water molecule: First, in that they are more fundamental than the “molecule” of a moral judgment – Moral judgment cannot exist apart from these elements, but the elements can and do exist without moral judgment; Second, and relatedly, in that these elements are not inherently tied to morality – Just as hydrogen and oxygen may contribute to various non-water molecules, the atomic elements of moral cognition may inform various nonmoral mental operations; however, when they come together in the right proportions, the proposed elements give rise to the “molecular” structure of moral cognition – a phenomenon that, much like H₂O, appears to be commonplace in nature.

Defining moral cognition in terms of nonmoral elements enables us to overcome two endemic problems with current definitions: First, they are often “question begging,” or logically circular. Sometimes this circularity is subtle. For instance, I show below that definitions of moral cognition that rely on an appeal to specific moral values (harm, justice, etc.) are circular because, to the extent that the appealed-to value is supposed to be thematically moral, it already contains within it an implicit idea of morality. It is this implicit idea of morality that needs to be explicitly rendered. More obviously, definitions that appeal to “moral goodness” are clearly circular in that the “moral” part of “moral goodness” is precisely what needs to be understood. The same holds for descriptions of “moral judgments” that do not interrogate what is distinctively moral about such judgments. The present approach overcomes such circularity by identifying the nonmoral elements of moral cognition. These “atomic” elements do not contain within themselves any implicit idea of morality. Nevertheless, the right combination of the right nonmoral elements in the right proportions give rise to an emergent property of moral cognition.

The field’s reliance upon question-begging operationalizations may be an attempt to find a way around another problem that has seemed intractable. To this day, debate persists over the extent to which morality and moral cognition imply the activity of a higher-order reflective capacity, such as “reason,” “rational deliberation,” or “metacognition.” Some adopt a narrow definition that includes a mandatory “reflection” criterion, whereas others prefer a broader definition that encompasses both reasoned moral cognition and implicit/automatic forms of “moral cognition,” such as “moral emotions” and “moral intuitions.” As has been argued (Fitzpatrick, 2017), the decision about whether to use the word “moral” in the broad or narrow sense is a matter of personal preference and is not addressable by current

evidence. Nevertheless, researchers' personal preferences have often featured centrally in scholarly debates over whether human infants or nonhuman animals possess moral cognition, making the question feel as unanswerable as the arguments are unresolvable.

The "molecular" definition I offer is not an opinion within the described debates but, instead, an end-run around them. The moral molecule is expansive, including implicit cognition, but those who prefer a narrower definition can add a "reflection" criterion without destroying the proposed molecular structure. Analogously, when one uses the word "water" in a way that distinguishes it from ice and water vapor, one does not thereby abandon the core structure of H_2O , but instead adds an additional criterion (liquidity) to the molecular definition of water. The molecular definition thus allows us to move beyond semantic debates over whether water *really is* a liquid or, instead, a broader category that includes solid and gaseous forms as well. Similarly, rather than becoming entangled in old arguments over the semantic meaning of "morality," researchers can move past these debates by clarifying their criteria up front. Are they seeking evidence of the core molecular structure defined in this article? Or are they looking for a rarer form of moral cognition, which includes both the described "moral molecule" and a capacity for rational reflection?

I will seek the molecular structure of moral cognition by identifying the conditions of possibility for a simple moral judgment. Toward this end, I must begin with a definite, if still underdeveloped, outline of what moral judgment is. Current operationalizations are inadequate for this purpose and typically suffer from the circularity problem described above. For instance, a recent review (Malle, 2021) organizes four distinct classes of "moral judgments" into a hierarchical taxonomy, moving from simple evaluations to more complex blame judgments. However, for each category of judgment within this taxonomy the question is not even posed – How does one distinguish a moral form of this judgment from nonmoral forms? Consider the category "evaluations." One may evaluate a pair of shoes, one's social behavior, or another's golf swing. But if one asks how to determine if any of these evaluations is a moral evaluation, moral psychology currently has no good answer. We merely assume certain evaluations are thematically moral, and then place these evaluations together into a common category, without offering a rigorous explanation for what makes them uniquely moral. The same considerations hold for Malle's second category: "norm judgments." Later in this article, I analyze the classic distinction between moral norms and nonmoral conventions as operationalized by Social Domain Theory (SDT). For now, I offer that there is currently contention over the precision with which this operationalization defines the moral domain, and that this contention emerges from the fact that we have not identified the nonmoral elements of a moral judgment. Finally, because each

judgment category in Malle's taxonomy builds upon those before, both "wrongness judgments" and "blame judgments" suffer from a similar conceptual ambiguity: It is not clear how to distinguish moral from nonmoral wrongness or moral from nonmoral blame.

My initial definition of a simple moral judgment will identify what is already implicitly part of the structure and indicate what still needs to be discovered. It is a straightforward analytic truth that moral judgments are normative, implying some sense of obligation to be or think or behave in one way rather than another. This axiom can be inferred from the fact that the polarities along which moral judgments are made – moral/immoral, right/wrong, good/evil – are all archetypally normative: one *ought* to be good and not evil, to do what is right and not what is wrong, to choose the moral over the immoral. As we have already seen, this criterion does not distinguish moral normativity from nonmoral normativity; nevertheless, the normative structure of moral judgment will prove to be an essential part of the definition of moral cognition.

It can also be deduced that moral judgments are relational. Specifically, moral judgments concern a relation between one or more moral subjects (i.e., those who are subject to moral norms) and one or more moral objects (i.e., that for the sake of which moral norms exist). This feature of relationality does not distinguish moral relationships from nonmoral relationships, yet it is still important to appreciate that moral judgment has this relational subject-object structure. I only caution that the described ideas of "subject" and "object" must be understood in a thoroughly phenomenological sense, not as necessarily discrete physical entities but, rather, as discrete interpretations or "framings," which may be imposed on any number of things. As such, it is even possible for the "same" entity to be understood as both subject and object of moral judgment—as in theories of virtue ethics (e.g., Aristotle, 1934), which propose that the subject who is morally obligated to cultivate virtue is also the moral object, whose happiness and well-being depends on their virtuousness.

The deduced features of normativity and relationality are not unique to the moral domain, but the present work aims precisely to identify what *is* distinctive about morality. That is, I propose to identify the emergent "molecular" structure that is specific to moral cognition and nothing else. Toward this end, I begin with a skeletal definition of moral judgment: Moral judgments are evaluations of (normative) moral obligations between moral subjects and moral objects. Obviously, this definition is circular, in that it includes within it an idea of morality that has yet to be articulated. However, this circularity will be broken when we have identified the nonmoral "atoms" that come together to give rise to the emergent "molecular" structure of moral cognition.

One final caveat is worth stating before I begin. The described moral molecule should be distinguished from explicit instances of moral approval or condemnation. Before one can morally disapprove of a particular attitude or behavior, one must already carry an implicit moral expectation that the attitude/behavior violates. This implicit moral expectation is already characterized by the described molecular structure, even before one makes any discrete “moral judgments.” Thus, although I aim to establish the molecular structure of a “simple” moral judgment, this structure is not identical with any particular judgment but is instead a horizon of possibility for any and all such judgments.

Within moral psychology, there are two research agendas that should be affected by the discovery of a moral molecule. First, there is the ultimate goal of modeling the complex dynamics of moral cognition. Elsewhere (Beal, 2020; Beal & Gogia, 2021), I describe some of the problems that emerge from the field’s lack of understanding of the deeper structure of moral cognition in contemporary models such as Moral Foundations Theory (MFT) (e.g., Graham et al., 2013, 2018; Haidt, 2012) and the Theory of Dyadic Morality (TDM) (e.g., Gray et al., 2012; Schein & Gray, 2015, 2018), while also analyzing the literature on moral dilemmas (e.g., Foot, 1967; Greene, 2013; Kohlberg, 1981; Mikhail, 2007; Thomson, 1985), dehumanization (e.g., Harris & Fiske, 2006, 2011; Haslam, 2006; Haslam & Loughnan, 2014; J.P. Leyens et al., 2007; J-P. Leyens et al., 2003), and moral expansiveness (Crimston et al., 2016, 2018). In the present article, I do not repeat this analysis of current models of moral cognition. Those who would like to understand the difference between the current approach and other models, such as TDM, may read the aforementioned articles.

The second research agenda, which I do engage here, is much simpler. In some contexts, researchers seek to determine whether or not any form of moral cognition is present at all. For instance, developmentalists studying the early emergence of moral cognition are interested in the question of when a simple form of moral cognition first emerges in human development. In a parallel line of research, ethologists (i.e., researchers studying animal behavior) are interested in knowing whether (and which) nonhuman animals possess any form of moral cognition at all. Clearly, the success of both research programs depends on the ability to identify what counts as moral cognition. In the absence of such a definition, it is difficult to distinguish good evidence of moral cognition (Case #1), whether in human infants or nonhuman populations, from false positives (Case #2). The present work focuses on the literature on early human development, showing that current definitions are insufficient, and I also argue that ethological research suffers from a similar conceptual ambiguity surrounding the idea of “moral cognition.” [Section 2](#) offers a critical analysis of the five most common approaches

to defining moral cognition, which are currently used by both developmentalists and ethologists. After engaging in this critical analysis of the current literature, I move on to identify the molecular structure of moral cognition in [Section 3](#).

2. Critical analysis of current definitions

2.1. *“Functionalist” definitions: Cooperation is not necessarily evidence of moral cognition*

Some developmentalists (e.g., Cowell & Decety, 2015; Hamlin, 2013, 2014; Hamlin & Van De Vondervoort, 2018; Tomasello, 2018) and ethologists (e.g., Bekoff & Pierce, 2009; Tomasello & Vaish, 2013) look to “functionalist” definitions emphasizing the social and evolutionary utility of morality for promoting cooperation (e.g., Curry, 2016; Graham et al., 2013; Greene, 2015; Haidt, 2012). According to this view, cooperation is the ultimate evolutionary explanation for morality, and it can therefore be said to define morality and moral cognition. Some of the theorists cited above even appear take this hypothesis a step further, reaching a conclusion about the “proximate” function of morality (i.e., its function in the present) and suggesting that morality can be defined, whether in the past or the present, by the function of promoting cooperation.

To understand why a “functionalist” approach is insufficient, it is useful to appreciate the classic distinction between ultimate (i.e., evolutionary) and proximate (e.g., psychological) levels of explanation: Because the function of any feature covaries with a changing ecological context, describing the reasons why a feature has evolved is not equivalent to explaining what the feature does here and now (see Mayr, 1961; Tinbergen, 1963; De Waal, 2008). Our question as to the nature of moral cognition is explicitly posed at the proximate level: Regardless of how moral cognition may have evolved, the question is how do we identify, in a precise and unambiguous way, what moral cognition is right now?

My intent here is not to criticize the above evolutionary approaches, but to identify several reasons why our proximate psychological question cannot be answered by pointing to the function of promoting cooperation: 1. Most obviously, nonmoral features may also promote cooperation: e.g., the capacity for language, the tendency to imitate others (e.g., Henrich & Gil-White, 2001), or even enlightened self-interest (e.g., Smith, [1776] 1904; Trivers, 1971). Thus, a “functionalist” definition would have to include additional criteria that could distinguish moral cooperation from nonmoral cooperation. 2. However, even these additional criteria would be insufficient, since it is also the case that morality sometimes *impedes* cooperation – e.g., the very theories that posit that morality functions to promote interindividual

cooperation also suggest that morality supports intergroup *competition* (e.g., Darwin, 1871; Greene, 2013; Haidt, 2012; Sober & Wilson, 1998). 3. Finally, posing the question at a proximate level of analysis leaves us with no good reason to deny that moral behaviors might at times promote neither cooperation nor competition – as may often be the case with private moral acts. As such, we cannot infer from any cooperative, uncooperative, or noncooperative behavior that an individual is employing moral cognition or not. In order to draw any such inference, we would have to begin by specifying, in the first place, what counts as moral cognition – but this is precisely the question before us.

2.2. Normative definitions: Moral goodness is not necessarily evidence of moral cognition

Some developmentalists appeal explicitly to the idea of moral goodness (e.g., Hamlin, 2013; Kohlberg, 1981) as a criterion for morality, while many others implicitly invoke this idea (e.g., Decety & Cowell, 2014; Gilligan, 1982; Piaget, 1965; Turiel, 2018). The ethological literature is also filled with anecdotes about virtuous behavior in nonhuman animals (see, e.g., Bekoff & Pierce, 2009; De Waal, 2008, 2009), which is sometimes interpreted as evidence of moral cognition in these species. Such accounts may be incredibly compelling; however, we must bear in mind that one may behave in ways that appear morally good without using moral cognition (i.e., Case #2). For instance, one might behave pro-socially out of calculated self-interest, or one might follow moral rules accidentally. Conversely, one might use moral cognition while behaving in ways that they themselves consider morally bad – as when one feels a pang of conscience while committing a “sin.” Finally, since what is viewed as morally good by some might be considered morally bad by others, humans and nonhumans may have moral reasons or motives for doing things that others perceive as immoral. Thus, just as seemingly good behavior is not unambiguous evidence of moral cognition, seemingly bad behavior is not necessarily evidence of an absence of moral cognition. To resolve the ambiguity, we must first define what moral cognition is.

2.3. Social norms: Normative social judgments are not necessarily evidence of moral cognition

Developmentalists sometimes describe the ability to distinguish between good and bad behaviors and dispositions as a “moral sense” (e.g., Hamlin, 2013; McAuliffe et al., 2019; Taborda-Osorio et al., 2019). However, humans and other species may sometimes have nonmoral reasons for making normative social judgments of behavior and character. For instance, one

might prefer helpful to unhelpful people for purely self-interested reasons (i.e., Case #2). Thus, the fact that members of a population are making normative social evaluations does not mean they are employing moral cognition to do so. The most fundamental questions remain: Within the broader genus of evaluations, how do we distinguish moral from nonmoral evaluations? Within the broader genus of norms, how do we distinguish moral from nonmoral norms? How can we tell if a normative social evaluation – whether of behavior, attitude, or character – is moral or not? To make this distinction, we need a more precise definition of moral cognition.

2.4. Moral values: Engaging with “moral” content is not necessarily evidence of moral cognition

Developmentalists and ethologists often appeal to common moral themes or values, such as harm, welfare, fairness, equitable distribution, and rights, and suggest that evidence of engagement with such themes might constitute evidence of moral cognition. For instance, Piaget (1965) thought moral cognition was essentially defined by a concern with fairness between people; Kohlberg defined moral cognition as “justice reasoning” (Kohlberg, 1984, pp. 215–217); Gilligan (1982) expanded the moral domain to include both justice and care for others; and Turiel (1983) defined the moral domain as involving consideration of justice, welfare, and rights between people. Ethologists Bekoff and Pierce (2009) describe concern with well-being and harm as defining features of morality, and De Waal (2006, p. 162) similarly defines the moral domain in terms of the “two H’s” of helping and harming. Looking at the current literature, I find that moral psychologists continue to define morality in terms of common moral themes, as evidenced by the fact that all the contributors to *Human Development’s* (2018) special issue on moral development attempt to define morality at least partially in terms of one or more of the themes or values listed above (Dahl, 2018; Hamlin & Van de Vondervoort, 2018; Killen et al., 2018; J. Smetana et al., 2018; Tomasello, 2018; Turiel, 2018). Similarly, although not all the articles in *Frontiers in Psychology’s* (Dawkins et al., 2019) recent special issue on the same topic offer an explicit definition of morality or moral cognition, those that do include the above moral themes/values in the definition (Dahl & Killen, 2018; Dawkins et al., 2019; Lee et al., 2018; Misch et al., 2018). Finally, outside the literatures on human development and animal behavior, an old debate between “monists” (e.g., Gray, 2014; Gray & Keeney, 2015a, 2015b; Gray et al., 2014, 2012; Royzman et al., 2015; Schein & Gray, 2015, 2018) and “pluralists” (e.g., Graham, 2015; Graham et al., 2013, 2018, 2011; Haidt, 2012; Haidt & Bjorklund, 2008; Haidt et al., 2015; Rottman et al., 2014) continues over the question of how many distinct values constitute the

“irreducible basic elements” of the moral sphere (see Beal, 2020 for a critical review).

Taking this approach to defining morality or moral cognition is circular, however. To define morality by appealing to specific thematic content begs the question of what makes a concern with justice, welfare, etc. *moral*. Without answering this deeper question, we cannot know how many values or themes might contain moral content, and we cannot know how to distinguish a properly moral engagement with “moral” themes from an engagement that does not employ moral cognition at all (i.e., Case #2). Consider that we can program machines like self-driving cars to navigate issues of harm, welfare, fairness, and rights, and yet we do not need to endow these machines with a special capacity for moral cognition. Similarly, even a person who lacked any capacity for moral cognition might think about harm from a medical perspective or reason about equitable distribution from a legal perspective or demand to have their political rights protected, and so on. Thus, it is possible to deal with these themes without making any moral judgments and without using moral cognition.

These considerations suggest that, instead of merely establishing that someone is thinking about or navigating the above themes, we need to identify what constitutes a moral as opposed to a nonmoral way of engaging these themes. That is, we need unambiguous criteria for what counts as moral cognition.

2.5. Social Domain Theory: Is distinguishing moral rules from conventions evidence of moral cognition?

Social Domain Theory (SDT) suggests that, across diverse societies, moral rules are characteristically understood as being more universal and context-independent (and, less consistently, more serious and punishable) than nonmoral “conventions.” Since Turiel’s early work, SDT theorists have observed this tendency to distinguish between moral rules and social conventions in both adults and children in diverse cultures across many studies (see reviews from J. G. Smetana et al., 2014; Nucci, 2001; Smetana, 2006; Tisak, 1995; Turiel, 1998, 2002, 2006), though there are gray areas between moral and conventional domains, such as moral “norms with a conventional element” (Heath, 2017). Today, many developmentalists, ethologists, and other researchers across the interdisciplinary field of moral psychology appeal to SDT’s distinction between moral and conventional domains when defining morality and moral cognition.

It is plausible that people, including young children, who spontaneously make this distinction are using moral cognition to do so. After all, SDT’s operationalization does more than merely show that children make the same kinds of distinctions that adults do. Evidently, children make the distinction in a similar way as adults (i.e., perceiving moral rules as universal and context-independent and conventions as relatively subjective and

situational). Moreover, the fact that people from diverse cultures and from an early age appear to make the same kind of distinction in the same way seems to suggest that humans have an innate tendency to make moral judgments (see, e.g., Joyce, 2006).

Despite these promising features of SDT criteria, the claim that they are evidence of a distinctively moral cognitive domain is controversial. For instance, Stich (2018) argues that SDT criteria do not constitute a cluster of naturally linked properties, and he cites evidence that the criteria of universality and context-independence are sometimes unlinked from the other SDT criteria (e.g., Fessler et al., 2015; Haidt et al., 1993; Kelly et al., 2007; Nichols, 2002; Nisan, 1987), a fact that would seem to rule out the possibility that these criteria straightforwardly define the moral domain.

3. The moral molecule

3.1. Overview and context

Stich (2018) concludes his critique of the evidence from SDT by ruling that moral judgment is probably not a “natural kind” corresponding to a distinctive area of cognitive performance. Fitzpatrick (2017) makes a similar argument when analyzing the literature on moral cognition in nonhuman animals, pointing out that different theorists are working with different preconceptions about what counts as “morality” and suggesting that there is no “correct” definition that could resolve such debates because, once again, moral cognition is not a natural kind. Although Stich and Fitzpatrick make apt criticisms of current approaches, I argue below that moral cognition is indeed a “natural kind” in the sense that it is a distinctive cognitive mode that naturally emerges as soon as an identifiable set of nonmoral cognitive operations are performed together; moreover, the described “moral molecule” is widespread and characteristic of all forms of recognizable moral judgment across diverse ages and cultures. As indicated above, all sides of the described “semantic” debates over what constitutes the proper domain of morality presuppose *at least* the “atomic” conditions I identify below. Thus, in Section 3, I identify these “atomic” nonmoral operations and show how they combine to form the unique “molecular” structure of moral cognition.

3.2.1. Moral objects and perceived inherent value (PIV)

Recall from Section 1 that the definition of a simple moral judgment depends on the definition of morality. As such, it would seem natural to begin the search for the meaning of morality by looking at the theories of moral metaphysics from which the meaning of morality historically derives. Is there any feature of morality upon which all theories agree? In fact, every

metaphysical theory of morality or ethics is grounded in some claim about what is inherently valuable. There are finer distinctions drawn in this space, between ideas of “the good” (Plato, 2004), “eudaimonia” (Aristotle. *The Nicomachean Ethics*, 1934), “pleasure” (Bentham, 2010), “intrinsic value” (e.g., Korsgaard, 1983; Moore, 1903), “good-in-itself” (Kant, 2002), “primary goods” (Rawls, 1971), and so on. Yet, these concepts are similar in that they all invoke an idea of value that is not merely relative to subjective opinion, socio-historical context, or instrumental usefulness – value that is, as such, supposed to inhere in things. Moreover, these moral philosophers all implicitly or explicitly distinguish their favored category of inherently valuable things from forms of noninherent value that are supposed to be, when taken on their own, irrelevant to morality: e.g., “instrumental value” or “relative ends.” Thus, although philosophers disagree over which particular form of inherent value is the proper basis for a prescriptive moral metaphysics, they evidently agree that one’s view of what is inherently valuable is essential for distinguishing moral right from wrong. I take from this analysis the thesis that perceived inherent value (PIV) is an implicit condition of “morality” in the discourse of moral philosophy, which is defining for moral objects.

Of course, to identify a common feature of philosophical understandings of morality is not the same as to show that moral cognition, defined in terms of this feature, is a “natural kind.” But what if this very same feature is a distinctive part of moral judgments made by people of diverse ages and cultures, including people who are unfamiliar with the history of Western philosophy and who do not know the meaning of the English word “morality”? Such a discovery would lend credibility to the argument that moral cognition is an identifiable natural kind, a distinctive and naturally emerging cognitive mode. In fact, I believe a good deal of evidence for this claim already exists and is simply awaiting the right interpretation. Toward this end, I turn back to the most well-studied operationalization of a moral judgment – that of Social Domain Theory (SDT).

In Section 2.5, I acknowledge Stich’s (2018) compelling argument that SDT criteria do not straightforwardly define the moral domain, since there is evidence that the criteria of universality and context-independence are sometimes unlinked from the other SDT criteria. However, there is another way to interpret the evidence from SDT. These criteria may instead be understood as “symptoms” of moral cognition. Diagnosing from symptoms involves considering the relevant context and determining the underlying condition(s). Once we have identified the more basic condition(s), variation at the level of symptoms – such as that observed by Stich and colleagues – can be explained.

Taking this diagnostic approach, I note that, although some of the SDT criteria are ambiguous, there are two criteria that, when understood as

symptoms, point unambiguously in one direction. If one is obligated to follow a rule that is supposed to be valid regardless of context (i.e., it is “context-independent”) and subjective opinion (i.e., it is “universal”), the logical implication is that this obligation must be grounded in value attributed to things themselves – value that makes these objects worthy (in perception) of moral consideration. As such, the evidence from SDT suggests that PIV is an implicit condition of moral judgment. Thus, SDT appears to confirm that, not only in philosophical discourse but also in the everyday lives of diverse people, PIV is an essential criterion that is defining for moral judgment. Notably, this central role of PIV does not seem to depend on one’s conceptual or theoretical understanding of the historical concept “morality” – since it is implicitly evoked by little children and non-Western populations. As such, the role of PIV within archetypal moral judgments does not appear to be a matter of mere semantics, but is more plausibly explained as a naturally emerging property of human perception, which is in fact utterly commonplace – a natural kind.

3.2.2. *Moral visibility and the status of moral objecthood*

Although the above considerations suggest that PIV is necessary for moral cognition, reflection shows that PIV is neither sufficient to define moral objects nor to give rise to moral judgments. Just because one perceives an object as inherently valuable does not mean one feels morally obligated toward the object. There may be many cases where inherent value is perceived, and yet no moral obligation is registered. For instance, one may feel that small bugs have some amount of inherent value, inasmuch as they are not mere instruments for one’s personal use but have their own lives and, plausibly, their own subjective experiences. Yet, at the same time, one may not feel moral obligations toward little bugs. Many philosophers have taken analogous observations to imply that there must be a special, uniquely moral category of PIV objects (PIVm) (see, e.g., Bentham, 2010; Kant, 2002; Korsgaard, 1983; Scanlon, 1998), which can be categorically distinguished from nonmoral PIV objects (PIV~m). According to this way of thinking, little bugs might seem to be PIV objects that fall outside the category of moral objects (i.e., they are PIV~m), whereas other objects, such as human beings, might seem to be categorically PIVm.

It is at least methodologically acceptable for the moral metaphysicians cited above to posit such categorical distinctions – they expressly aim to determine what *really is* worthy of moral consideration, regardless of how various people happen to see things. However, moral psychology has a different aim. When perceptions diverge, a moral psychologist’s job is not to say who is ultimately correct but, rather, to explain the psychological source of the difference. As such, within the discipline of moral psychology there is a problem with taking the described “categorical” approach to

defining moral objects in cognition. Because perceptions of what is truly a moral object vary, not only between populations and individuals but even within the same individuals over time, it is always possible to falsify any particular hypothesis about a categorical PIVm in cognition. All one has to do is identify *any* PIV object outside the proposed category that *some* moral subjects view as moral objects. No hypothesized category withstands scrutiny. There are people who do not feel moral obligations toward little bugs, but there are also people who do feel moral obligations toward little bugs. Moreover, the same person might feel no moral obligations toward little bugs in one context, but then feel moral obligations toward little bugs in another context (see Beal & Gogia, 2021).

Thus, moral psychologists face a difficult problem. Because PIV is insufficient to call forth a moral obligation in every case, many have assumed that there must be a special category of moral objects (PIVm) toward which people feel moral obligations. However, the fact that people's moral objects differ from those of other people, or even from themselves at different moments, means that, within cognition, an objective-categorical PIVm does not exist. This absence of an objective category of moral objects is an obstacle because our "shell" definition in [Section 1](#) indicates that we need a definition of moral objects, in order to define moral judgments.

The path forward involves abandoning the hypothesis of an objective-categorical PIVm and understanding that PIVm is a *status* that PIV objects achieve in cognition when certain conditions have been met – Let us call this status "moral visibility." An object becomes morally visible if and when it meets the relevant conditions in cognition to be perceived as a moral object (PIVm). My hypothesis, which is grounded in both intuition and burgeoning data, is that moral visibility is primarily a function of how much inherent value the subject attributes to the object and how "close" the subject feels to the object. If the object is perceived as possessing a sufficient amount of inherent value and being sufficiently "close" to the subject, then the subject will feel a sense of moral responsibility toward the object – i.e., The object will become morally visible and achieve the status of PIVm. However, if the object does not reach the threshold for moral visibility in a subject's mind – whether because of insufficient perceived value, insufficient perceived "closeness," or a combination of the two – no sense of moral responsibility will emerge. The object will remain morally invisible.

The proposed criteria must be understood with the proper nuance. "Closeness" should here be interpreted not in an objective physical sense but in an existential sense of intimacy/identity¹ (cf. Dasein's existential "spatiality" in Heidegger, 2010). I use the expression "existential closeness" rather than the more common "social closeness" (e.g., Jones & Rachlin, 2006; Law et al., 2021) because I want to accommodate moral relationships that may not be intuitively described as "social" (e.g., moral relationships

with things, places, and ideas). However, these constructs are closely related, and social closeness may be considered a centrally important type of the broader genus, “existential closeness.”

The idea that moral visibility is defined by the amount of perceived inherent value and one’s perceived existential closeness to the inherently valuable object is currently just a hypothesis. But this hypothesis is consistent with recent research into “moral circles” (e.g., Crimston et al., 2016, 2018), “dehumanization” (e.g., Haslam, 2006; Haslam & Loughnan, 2014), and “biased benevolence,” (Law et al., 2021). Together, these research agendas suggest that perceived closeness (e.g., being seen as more or less human) and/or perceived value (e.g., being seen as a more or less valuable human) shapes the amount and nature of our perceived moral responsibilities toward various objects. The cited data do not prove my claim, but they are strikingly consistent with it. It should be possible to empirically investigate precisely what contributes to an object’s moral visibility, and I propose ways of exploring such hypotheses elsewhere (Beal & Gogia, 2021). For now, I would like to make a more basic point: The generic proposition that a simple moral judgment relies upon *some* conditional moral visibility is more than a mere conjecture. As long as the contrary claim of an objective category of moral objects is false – and the above argument shows that it is false – the proposition that moral visibility is a subjective-conditional status must be true. We can debate precisely what dimensions contribute to moral visibility, but such a debate presupposes the truth of the generic “moral visibility” proposition.

Thus, we have defined the conditions for moral objecthood: The object must be perceived as inherently valuable (PIV), and it must meet the conditions for moral visibility (PIVm).

3.2.3. *Moral subjects and perceived agency (PA)*

So far, we have looked at just one side of the primordial moral relationship – moral objects. But moral norms are ultimately directed toward moral subjects in their relations to moral objects. Accordingly, in this section, I will look at the conditions underlying perceived moral subjecthood.

In order to perceive oneself or another as being obligated to conform to a moral norm, whether such a perception is conscious and reflective or unconscious and implicit, one must perceive oneself or another as being agentive. Perception of agency (PA) is thus another essential condition of moral judgment (and also of nonmoral obligations).²

Unlike PIV, which we arrived at inductively by looking at empirical evidence, the PA criterion is discovered through straightforward deduction. It is an “analytic proposition,” which is contained within the concept of “obligation.” This fact means that, as long as our shell definition from

Section 1 is accurate, we can be confident that some form of PA is a necessary condition of moral cognition.

It is important, however, to form a clear picture of precisely what form of perceived “agency” is presupposed by moral judgment. Agency is sometimes understood in a specialized sense of a “free will,” which implies an ability to choose that is independent of deterministic causal forces. However, agency may also be understood in a broader sense of one’s positive capabilities to accomplish, or to try to accomplish, some moral end. This broader sense of agency may involve perceived free will, but it may also involve other capabilities that make moral actions or moral goodness possible (cf. “capacitarian” theories of agency, as described by Murray, 2018).

The broad sense of agency-as-capability is important for understanding the relation between PA and moral character traits. A character trait is distinctively moral to the extent that it inclines one toward moral behavior. One may be inclined to moral behavior by one’s will, but other features matter as well. For instance, in *The Nicomachean Ethics* (1934), Aristotle imagined that one might become so virtuous that one would consistently behave in the right manner out of pure habit, without deploying any effortful act of will. Acting out of one’s virtuous moral character still involves a kind of agency, in that one would not have the power to behave in a consistently virtuous manner if one did not possess these character traits. Yet, this form of agency is not reducible to an immediate exercise of voluntaristic will.

The broad concept of “agency” described above also helps us to properly interpret evidence that different cultures vary in how much they emphasize the role of intentions when making moral judgments (see Barrett et al., 2016), as well as the fact that, across cultures and individuals, moral blame is about more than just our judgment of intentions – consequences also matter (see Martin & Cushman, 2016; Patil et al., 2017). Evidence that some populations deemphasize the moral significance of intentions relative to other populations is not evidence against the PA criterion. Regardless of the extent to which a culture or individual perceives moral subjects to be exercising a voluntaristic “free will,” any moral judgment depends upon the perception of some form of agency (PA).

3.2.4. Moral vision and moral agency

If PA and PIVm are indeed necessary criteria for moral cognition, we must ask once again whether these criteria are sufficient to characterize the molecular structure of moral judgment. Once again, it appears that our criteria are insufficient. Consider the following counterfactual scenario. Suppose my beloved child is swimming in the ocean when a shark attacks her. In my opinion, has the shark done something immoral? I am ready to admit that the shark has exerted some form of agency in biting my child. Specifically, it

seems to me that the shark attacked my child because it wanted or intended to, in some minimal sense of these terms – e.g., I would distinguish this action from that of a remotely controlled robot shark, whose movements I would attribute instead to the person with the remote. Hence, for me, the condition PA is met in some minimal sense. In addition, it seems to me that my child has a tremendous amount of inherent value and is, in my mind, a moral object. Thus, in my mind the criteria PA and PIVm are met. Yet, I would still not say that the shark has behaved immorally. Some unspecified condition, which is a necessary part of moral judgment, has not been met in my mind, and we must figure out what that condition is.

Just as it was necessary to nuance the criterion of PIV by introducing the conditional status of moral visibility, it is also necessary to nuance the PA criterion with a condition that I will call “moral vision.” Although I perceive the shark as an organism that acts intentionally, and in this sense exercises agency, I do not expect the shark to see my child as a moral object (PIVm). That is, in its relation to my child, I do not think the shark possesses moral vision. Thus, although the shark possesses other perceptive faculties and can use them to act intentionally, in my mind it can have no awareness of my child as a moral object, and the shark therefore cannot exercise moral agency toward my child. Thus, in my view, while the event is certainly tragic, there is no sense in morally blaming the shark. On the other hand, if a friend were to attack my child, I would register this behavior as immoral because I do attribute moral vision to my friend. That is, I do expect my friend to be capable of seeing my child as a moral object and understanding that attacking my child is immoral. The moral agency I attribute to my friend (but not to the shark) emerges from the combination of agency and moral vision that I perceive my friend to possess. These conditions are together defining for perceived moral agency (PAm), or moral subjecthood.

3.2.5. *The “morality threshold” and the moral molecule*

We finally have the molecular structure of a simple moral judgment – the perception of normative obligation that emerges naturally from the perceived relationship between moral subjects and moral objects. Moral subjects are those who possess PAm, defined as perceived agency plus perceived moral vision. Moral objects are those that possess PIVm, defined as perceived inherent value plus moral visibility. All moral judgments presuppose the perception of a moral relationship that is defined by these basic elements. As such, moral cognition is an ability to see things in a moral way, which is a horizon of possibility for all discrete moral judgments.

Although I have proposed two distinct conditions in “moral visibility” and “moral vision,” these conditions are really two sides of the same coin. If one entirely lacks moral vision, all objects are morally invisible, and vice versa. Thus, these two conditions are distinct approaches to the same

threshold phenomenon, which is defining for moral relationships – Let us call it the “morality threshold” (MT). Just as the visibility of a physical object is a function of object-features (e.g., size), subject-features (e.g., vision), and relational features (e.g., physical closeness of the subject to the object), so the moral visibility of an object is a function of object-features (e.g., PIV amount), subject-features (e.g., moral vision), and relational features (e.g., existential closeness of the subject to the object). That is, before a moral relationship will be perceived, the object must be perceived as having enough inherent value to warrant consideration for its own sake (i.e., not *exclusively* for some other end); the subject must be perceived as possessing adequate moral vision to see PIVm objects and adequate agency to treat the object as it deserves; and the object must be perceived as being existentially close enough to the subject to warrant moral consideration. I propose that these factors conspire to define a threshold (MT) below which no moral relationship – and accordingly, no moral obligations – will be perceived. And I reiterate that, although the precise nature of MT is a matter for empirical investigation, the above argument suggests that some threshold for moral visibility must exist, and that MT is a necessary condition for moral cognition.

MT thus defines the subscript m that is attached to both moral subjects and moral objects and expresses the reality that moral cognition is an emergent phenomenon. Synthesizing the “atomic” elements PIV, PA, and MT yields moral cognition – a distinctively moral mode of perception. This perceptual mode, “moral sight,” is the irreducible form of moral cognition – the “moral molecule.”

This is the crux of the argument: Like all molecules, the moral molecule is irreducible in the specific sense that any decomposition of the described structure necessarily eliminates the phenomenon. To say that an object has inherent value is not to make a moral judgment but simply to place the object within a particular category – PIV. PIV only becomes morally relevant when the object is perceived to be morally visible and to be in the presence of someone who is perceived as a moral agent. Similarly, the perception of agency is not, in itself, a moral judgment. PA only becomes morally relevant in relation to PIVm objects, at which point the subscript “m” may attach to the perceived agent as well. Just as the atomic elements hydrogen and oxygen are not water on their own, but when combined together in the right proportions they give rise to a simple water molecule, so the elements constitutive of PAm and PIVm are not moral on their own, but in combination they give rise to the “molecular” structure of moral cognition.

3.3. Nuances

3.3.1. *The moral molecule is an implicit horizon for all explicit moral judgments*

It is important to appreciate that, although the proposed structure is characteristic of all discrete moral judgments, it is actually more fundamental than these judgments themselves. What I am describing is a perceptual capacity, rather than any specific perception. Typically, our capacity for moral sight only becomes explicit in response to exceptional moral violations or good deeds. We take notice when someone who we perceive as a moral subject goes outside the (normative) lines with respect to something we (implicitly) perceive as a moral object, or when they go to remarkable lengths to do the morally right thing. In unusual circumstances such as this, we may experience moral emotions, such as blame or praise. It is true that these experiences of morally blaming or morally approving of a person or behavior are what is commonly understood by a “moral judgment.” However, it is important to appreciate that these discrete moments of moral blame or approbation presuppose that one has already been continuously perceiving a moral relationship with embedded moral expectations. This implicitly moral way of seeing already bears the molecular signature of moral cognition, even before one makes any explicit “moral judgments.” The proposed “moral molecule” may thus be understood as a cognitive horizon of possibility for all discrete moral judgments, moral emotions, moral intuitions, and so on.

3.3.2. *This molecular structure is both necessary and sufficient for the emergence of moral cognition*

Establishing necessary conditions for a phenomenon is much easier than establishing sufficient conditions, and it is worth asking whether the described elements of a moral relationship are sufficient, in themselves, to bring about or constitute moral cognition. Thus, in this section I consider whether there are any exceptions to the proposed moral molecule, and whether this molecule is, in itself, moral cognition.

One way to make this inquiry is to ask whether it is possible in principle for someone to see themselves as PAm and an object as PIVm, yet recognize no moral obligation toward the object. If my thesis is correct, this is categorically impossible. In the same way that H₂O just *is* water, so the perception of a moral relationship just *is* moral cognition – even in the absence of any explicit moral condemnation or approbation. Because I am making such a strong claim, I must be able to explain phenomena that, at first glance, look like falsifying exceptions to this rule.

I will consider three kinds of phenomena that might seem to falsify my claim. First, consider the kind of moral indifference observed in, for instance, psychopathy. It is possible for a moral subject to recognize that

an object is PIV_m in the eyes of others while remaining morally indifferent toward the object themselves. Does this moral indifference toward a moral object stand as a falsifying example that contradicts the above thesis? The following considerations show that the answer is “no.” If the question is whether the subject has moral cognition, it is not relevant whether the subject knows that the object is morally visible to others. Just as a blind person does not gain the power of sight through knowledge of what others see, a person who fails to see the inherent value of an object for themselves cannot gain moral sight of the object through knowledge of what others see. Utter moral indifference is only possible if, for this subject at this moment, the object is morally invisible.

Second, one might point out that, in fact, people are constantly indifferent toward things that they themselves implicitly consider moral objects. For instance, driving my car down the road, I may feel no particular sense of moral obligation toward strangers on the sidewalk; however, if you ask me whether these people are moral objects, I will assure you that they are. Is there a contradiction here? Once again, the answer is “no.” A simple thought experiment reveals that my seeming indifference toward these pedestrians is very far from the “utter indifference” of someone who lacks moral sight. Suppose I accidentally drive my car into one of these pedestrians. At this moment, the guilt that washes over me testifies to the moral relationship in which I must have implicitly understood myself to be. Only if I were able to run over pedestrians and feel no remorse at all would it be appropriate to infer that I lack moral sight with respect to the pedestrians – or, equivalently, that they are “morally invisible” to me.

Third, and finally, there may be special kinds of relationship that seem like exceptions to the described structure of moral sight. For instance, those who perceive themselves as owners of moral objects sometimes think they have a moral right to treat their own possessions however they see fit. This phenomenon might seem like a contradiction, since there would appear to be objects that a single subject sees as simultaneously PIV_m and PIV_{~m}. For instance, a person who considers themselves as PIV_m may nonetheless deny that they have moral obligations toward *their own* body, while still feeling that others ought to show them moral consideration.

This apparent contradiction is illusory, however. Taken to its logical extreme, perceived ownership undermines the inherence of perceived value, because the value one perceives in the object ultimately redounds to the object’s owner. Thus, if one considers oneself to be an *absolute and exclusive* owner over an object, one effectively erases, in one’s own mind, the possibility of being morally obligated toward that which one believes oneself to possess. By the same logic, to the extent that a possession is supposed to belong to the owner, mistreatment of the possession will be coded as mistreatment of the owner. Thus, one may simultaneously feel they have

a right to treat their own body as they like and also feel that others do not have such a right. This phenomenon emerges from the *interaction* of the moral relationship with the ownership relationship and does not conflict with the proposed structure of moral cognition.

The above considerations are consistent with the claim that there are no exceptions to the proposed “molecular” structure of moral cognition, and that the described elements are not only necessary but also sufficient to constitute a moral relationship. If an object is perceived as inherently deserving consideration, and a subject is perceived as being able to treat the object appropriately or inappropriately, it follows that the subject ought to treat the object as it deserves to be treated – and ought not to violate these implicit entailments of the moral relationship. This conclusion is nothing other than an elaboration of the distinctive molecular structure of moral cognition.

3.3.3. *The subject-object structure of moral cognition is folded into moral character perception*

The described “background” perception of moral relationships is requisite not only for the emergence of moral judgments but even for the perception of moral character traits – a thesis I will argue in four steps: Step 1. Both moral and nonmoral character traits are defined by the behaviors they are supposed to promote, on average and in the long run. For instance, competent people are supposed to behave competently, kind people are supposed to express kindness, and honest people are supposed to be honest. In any given situation, this relation between a character trait and a behavior may not hold: Competent people make mistakes; kind people can be mean; and honest people occasionally lie. But a person who never behaves competently is not competent; a person who rarely shows kindness is not kind; and a person who constantly lies is not honest. 2. What distinguishes moral from nonmoral character traits is that moral traits are specifically supposed to incline one to moral or immoral behavior. 3. As I have argued above, the perception that a behavior is moral or immoral is determined by the perception that certain “moral objects” call for specific forms of moral consideration from “moral subjects.” 4. Synthesizing steps 2 and 3, we can see that the perception of moral character traits, as described in point 2, depends on the perception of moral relationships, as described in point 3. This formal priority of perceiving moral relationships holds even if it is also true that moral judgments tend to be motivated by our central concern with others’ character (see Uhlmann et al., 2015).

3.3.4. *Moral objects and subjects may also be nonmoral objects and subjects*

There is no contradiction in observing that objects are complex. For instance, a person may be seen as a moral object, deserving of respect and,

at the same time, as a material resource. One may be perceived as both a human being and an employee, a person with a rich inner life and a sexual object. There may be competition between these distinct aspects of an object, and such competition does indeed potentiate conflict – for instance, when it comes to distinguishing exploitation from profitable working relationships. However, I note that there is no logical contradiction entailed by such conflict, merely a reality of complexity – The possession of moral sight does not preclude the possession of other forms of sight.

4. Conclusion

My critique of current definitions of moral cognition in [Section 2](#) focuses on the problem of false positives (though I also point to false negatives). Behaviors that we tend to interpret in a moral way may also indicate nonmoral cognitive processes – such as when people cooperate, exhibit “moral goodness,” navigate social norms, or deal in “moral” themes out of nothing more than enlightened self-interest.

My proposal offers a conceptual solution to such ambiguity. Key to distinguishing false positives from real evidence of moral cognition is to show that an observed behavior requires the observed to possess moral sight. If cooperation or other “morally good” behaviors are performed for moral reasons – i.e., because one feels obligated by a perceived moral relationship – then the behavior is evidence of moral cognition. Similarly, if one’s social judgments or abstract values derive from moral considerations – i.e., from consideration of one’s obligations within moral relationships – then these judgments or values are evidence of moral cognition. Moral judgments emerge from the (often implicit) perception of a moral relationship between perceived moral subjects (PAm) and moral objects (PIVm), each of which is composed of the more basic, nonmoral cognitive elements described above. Developmentalists, ethologists, and anyone else looking for evidence of moral cognition, whether in human infants, adults with psychological disorders (such as antisocial personality), nonhuman species, or artificial intelligence, must offer evidence that these conditions of “moral sight” are operative. It is as simple as that. But no simpler.

Notes

1. Although intimacy and identity are distinct representations, their moral function may be roughly the same. For instance, if one does not identify as being similar to an object, one may nonetheless develop intimacy with the object, and this intimacy can create the necessary closeness for a moral relationship. Conversely, if one is not intimate with an object, one’s sense of categorical closeness with the object (e.g.,

being human, being family, being one's pet) can create the necessary sense of closeness for moral obligation (e.g., "I owe it to them because they're human beings, just like me").

2. Note that this claim is not a metaphysical defense of "free will" but simply a logical analysis of the concept of obligation. Even if reality is purely deterministic, this does not change the fact that the idea of obligation entails the *perception* of agency.

Acknowledgement

I thank Patrick Bixby and an anonymous reviewer for thought-provoking comments on earlier drafts of this article.

Disclosure of potential conflicts of interest

No potential conflict of interest was reported by the author(s).

References

- Aristotle. (1934). *The Nicomachean Ethics*. E. Capps, T.E. Page, & W.H.D. Rouse. (Eds.) (H. Rackham, Trans.). William Heineman.
- Barrett, H., Bolyanatz, A., Crittenden, A., Fessler, D., Fitzpatrick, S., Gurven, M., Henrich, J., Kanovsky, M., Kushnick, G., Pisor, A., Scelza, B., Stich, S., von Rueden, C., Zhao, W., & Laurence, S. (2016). Intentions and moral judgment across societies. *Proceedings of the National Academy of Sciences*, 113(17), 4688–4693. <https://doi.org/10.1073/pnas.1522070113>
- Beal, B. (2020). What are the irreducible basic elements of morality? A critique of the debate over monism and pluralism in moral psychology. *Perspectives on Psychological Science*, 15(2), 273–290. <https://doi.org/10.1177/1745691619867106>
- Beal, B., & Gogia, G. (2021). Cognition in moral space: A minimal model. *Consciousness and Cognition*, 92, 103134. <https://doi.org/10.1016/j.concog.2021.103134>
- Bekoff, M., & Pierce, J. (2009). *Wild justice: The moral lives of animals*. The University of Chicago Press.
- Bentham, J. (2010). *An introduction to the principles of morals and legislation*. White Dog Publishing. (originally published in 1781).
- Cowell, J., & Decety, J. (2015). Precursors to morality in development as a complex interplay between neural, socioenvironmental, and behavioral facets. *PNAS*, 112(41), 12657–12662. <https://doi.org/10.1073/pnas.1508832112>
- Crimston, C., Bain, P. G., Hornsey, M. J., & Bastian, B. (2016). Moral expansiveness: Examining variability in the extension of the moral world. *Journal of Personality and Social Psychology*, 111(4), 636–653. <http://doi.org/10.1037/pspp0000086>
- Crimston, C., Hornsey, M. J., Bain, P. G., & Bastian, B. (2018). Toward a psychology of moral expansiveness. *Current Directions in Psychological Science*, 27(1), 14–19. <https://doi.org/10.1177/0963721417730888>
- Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. In T. K. Shackelford & R. D. Hansen (Eds.), *Evolutionary psychology. The evolution of morality* (pp. 27–51). Springer International Publishing. https://doi.org/10.1007/978-3-319-19671-8_2

- Dahl, A. (2018). New beginnings: An interactionist and constructivist approach to early moral development. *Human Development*, 61(4–5), 232–247. <https://doi.org/10.1159/000492801>
- Dahl, A., & Killen, M. (2018). A developmental perspective on the origins of morality in infancy and early childhood. *Frontiers in Psychology*, 9, 1736. <https://doi.org/10.3389/fpsyg.2018.01736>
- Darwin, C. (1871). *The descent of man and selection in relation to sex*. John Murray.
- Dawkins, M., Sloane, S., & Baillargeon, R. (2019). Do infants in the first year of life expect equal resource allocations? *Frontiers in Psychology*, 10, 116. <https://doi.org/10.3389/fpsyg.2019.00116>
- De Waal, F. (2006). *Primates and philosophers: How morality evolved*. (S. Macedo & J. Ober, Eds.). Princeton University Press.
- De Waal, F. (2008). Putting the altruism back in altruism: The evolution of empathy. *Annual Review of Psychology*, 59, 279–300. <https://doi.org/10.1146/annurev.psych.59.103006.093625>
- De Waal, F. (2009). *The age of empathy: Nature's lessons for a kinder society*. Harmony Books.
- Decety, J., & Cowell, J. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science*, 9(4), 525–537. <https://doi.org/10.1177/1745691614545130>
- Fessler, D., Barrett, H., Kanovsky, M., Stich, S., Holbrook, C., Henrich, J., Bolyanatz, A., Gervais, M., Gurven, M., Kushnick, G., Pisor, A., von Rueden, C., & Laurence, S. (2015). Moral parochialism and contextual contingency across seven disparate societies. *Proceedings of the Royal Society B*, 282(1813). <https://doi.org/10.1098/rspb.2015.0907>
- Fitzpatrick, S. (2017). Animal morality: What is the debate about? *Biological Philosophy*, 32, 1151–1183. <https://doi.org/10.1007/s10539-017-9599-6>
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5–15.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development* (1993 ed.). Harvard University Press.
- 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). Academic Press. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>
- Graham, J. (2015). Explaining away differences in moral judgment: Comment on Gray and Keeney (2015). *Social Psychological & Personality Science*, 6(8), 869–873. <https://doi.org/10.1177/1948550615592242>
- Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitch, 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). Guilford Press.
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2), 366–385. <https://doi.org/10.1037/a0021847>
- Gray, K. (2014). Harm concerns predict moral judgments of suicide: Comment on Rottman, Kelemen and Young (2014). *Cognition*, 133(1), 329–331. <https://doi.org/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(8), 874–877. <https://doi.org/10.1177/1948550615592243>

- 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(8), 859–868. <https://doi.org/10.1177/1948550615592241>
- 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(4), 1600–1615. <https://doi.org/10.1037/a0036149>
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23(2), 101–124. <https://doi.org/10.1080/1047840X.2012.651387>
- Greene. (2015). The rise of moral cognition. *Cognition*, 135, 39–42. <http://dx.doi.org/10.1016/j.cognition.2014.11.018>
- Greene, J. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin 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). MIT Press.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon. Knopf Doubleday Publishing Group.
- Haidt, J., Graham, J., & Ditto, P. (2015, October 28). *The Volkswagen of moral psychology [Blog post]*. <http://www.spsp.org/blog/dyadic-morality>
- 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(4), 613–628. <https://doi.org/10.1037/0022-3514.65.4.613>
- Hamlin, K. (2013). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science*, 22(3), 186–193. <https://doi.org/10.1177/0963721412470687>
- Hamlin, K. (2014). The origins of human morality: Complex socio-moral evaluations by preverbal infants. In J. Decety & Y. Christen (Eds.), *New frontiers in social neurosciences* (pp. 165–188). Springer. https://doi.org/10.1007/978-3-319-02904-7_10
- Hamlin, K., & Van de Vondervoort, J. (2018). Infants' and young children's preferences for prosocial over antisocial others. *Human Development*, 61(4–5), 214–231. <https://doi.org/10.1159/000492800>
- Harris, L., & Fiske, S. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, 17(10), 847–853. <https://doi.org/10.1111/j.1467-9280.2006.01793.x>
- Harris, L., & Fiske, S. (2011). Dehumanized perception: A psychological means to facilitate atrocities, torture, and genocide? *Zeitschrift fur Psychologie*, 219(3), 175–181. <https://doi.org/10.1027/2151-2604/a000065>
- Haslam, N. (2006). Dehumanization: An integrated review. *Personality and Social Psychology Review*, 10(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4
- Haslam, N., & Loughnan, S. (2014). Dehumanization and inhumanization. *Annual Review of Psychology*, 65(1), 399–423. <https://doi.org/10.1146/annurevpsych-010213-115045>
- Heath, J. (2017). Morality, convention and conventional morality. *Philosophical Explorations*, 20(3), 276–293. <https://doi.org/10.1080/13869795.2017.1362030>
- Heidegger, M. (2010). *Being and time*. (J. Stambaugh & D. Schmidt, Trans.). State University of New York Press. (Original work published 1927).
- Henrich, J., & Gil-White, F. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, 22(3), 165–196. [https://doi.org/10.1016/S1090-5138\(00\)00071-4](https://doi.org/10.1016/S1090-5138(00)00071-4)

- Jones, B., & Rachlin, H. (2006). Social discounting. *Psychological Science*, 17(4), 283–286. <https://doi.org/10.1111/j.1467-9280.2006.01699.x>
- Joyce, R. (2006). *The evolution of morality*. The MIT Press.
- Kant, I. (2002). *Groundwork for the metaphysics of morals*. A. Wood (Ed., Trans.). Yale University Press. (Original work published 1785).
- Kelly, D., Stich, S., Haley, K., Eng, S., & Fessler, D. (2007). Harm, affect and the moral/conventional distinction. *Mind and Language*, 22(2), 117–131. <https://doi.org/10.1111/j.1468-0017.2007.00302.x>
- Killen, M., Elenbaas, L., & Rizzo, M. (2018). Young children's ability to recognize and challenge unfair treatment of others in group contexts. *Human Development*, 61(4–5), 281–296. <https://doi.org/10.1159/000492804>
- Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of justice* (Vol. 1). Harper & Row.
- Kohlberg, L. (1984). *The psychology of moral development: The nature and validity of moral stages* (Vol. 2). Harper & Row.
- Korsgaard, C. (1983). Two distinctions in goodness. *The Philosophical Review*, 92(2), 169–195. <https://doi.org/10.2307/2184924>
- Law, K. F., Campbell, D., & Gaesser, B. (2021). Biased benevolence: The perceived morality of effective altruism across social distance. *Personality and Social Psychology Bulletin*. <https://doi.org/10.1177/01461672211002773>
- Lee, K., Esposito, G., & Setoh, P. (2018). Preschoolers favor their ingroup when resources are limited. *Frontiers in Psychology*, 9, 1752. <https://doi.org/10.3389/fpsyg.2018.01752>
- 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. <https://doi.org/10.1111/j.1751-2409.2007.00006.x>
- Leyens, J.-P., Cortes, B., Demoulin, S., Dovidio, J., Fiske, S., Gaunt, R., Paladino, M.-P., Rodriguez-Perez, A., Rodriguez-Torres, R., & Vaes, J. (2003). Emotional prejudice, essentialism, and nationalism The 2002 Tajfel lecture. *European Journal of Social Psychology*, 33(6), 703–717. <https://doi.org/10.1002/ejsp.170>
- Malle, B. (2021). Moral judgments. *Annual Review of Psychology*, 72(3), 1–3.26. <https://doi.org/10.1146/annurev-psych-072220-104358>
- Martin, J., & Cushman, F. (2016). Why we forgive what can't be controlled. *Cognition*, 147, 133–143. <https://doi.org/10.1016/j.cognition.2015.11.008>
- Mayr, E. (1961). Cause and effect in biology. *Science*, 134(3489), 1501–1506. <https://doi.org/10.1126/science.134.3489.1501>
- McAuliffe, K., Bogese, M., Chang, L., Andrews, C., Mayer, T., Faranda, A., Hamlin, J. K., & Santos, L. (2019). Do dogs prefer helpers in an infant-based social evaluation task? *Frontiers in Psychology*, 10, 591. <https://doi.org/10.3389/fpsyg.2019.00591>
- Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, 11(4), 143–152. <https://doi.org/10.1016/j.tics.2006.12.007>
- Misch, A., Over, H., & Carpenter, M. (2018). The whistleblower's dilemma in young children: When loyalty trumps other moral concerns. *Frontiers in Psychology*, 9, 250. <https://doi.org/10.3389/fpsyg.2018.00250>
- Moore, G. E. (1903). *Principia ethica*. Cambridge University Press.
- Murray, S. (2018). Why value values? *Behavioral and Brain Sciences*, 41, E54. <https://doi.org/10.1017/S0140525X17000784>
- Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, 84(2), 221–236. [https://doi.org/10.1016/S0010-0277\(02\)00048-3](https://doi.org/10.1016/S0010-0277(02)00048-3)
- Nisan, M. (1987). Moral norms and social conventions: A cross-cultural comparison. *Developmental Psychology*, 23(5), 719–725. <https://doi.org/10.1037/0012-1649.23.5.719>

- Nucci, L. (2001). *Education in the moral domain*. Cambridge University Press.
- Patil, I., Calò, M., Fornasier, F., Cushman, F., & Silani, G. (2017). The behavioral and neural basis of empathic blame. *Scientific Reports*, 7(1), 5200. <https://doi.org/10.1038/s41598-017-05299-9>
- Piaget, J. (1965). *The moral judgment of the child*. (M. Gabain, Trans.). The Free Press. (Original work published 1948).
- Plato. (2004). *'Protagoras' and 'Meno'*. (R. Bartlett, Trans.). Cornell University Press.
- Rawls, J. (1971). *A theory of justice*. Belknap Press of Harvard University Press.
- Rottman, J., Keleman, D., & Young, L. (2014). Purity matters more than harm in moral judgments of suicide: Response to Gray. *Cognition*, 133(1), 332–334. <https://doi.org/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.
- Scanlon, T. (1998). *What we owe to each other*. Belknap Press.
- 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(8), 1147–1163. <https://doi.org/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(1), 32–70. <https://doi.org/10.1177/10888683176982>
- Smetana, J., Jambon, M., & Ball, C. (2018). Normative changes and individual differences in early moral judgments: A constructivist developmental perspective. *Human Development*, 61(4–5), 264–280. <https://doi.org/10.1159/000492803>
- 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). 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). Taylor & Francis.
- Smith, A. (1904). *An inquiry into the nature and causes of the wealth of nations*. (E. Cannan Ed.). Methuen. in 2 vols. (Original work published 1776).
- Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Harvard University Press.
- Stich, S. (2018). The quest for the boundaries of morality. In K. Jones, M. Timmons, & A. Zimmerman (Eds.), *The routledge handbook of moral epistemology*. Routledge.
- Taborda-Osorio, H., Lyons, A., & Cheries, E. (2019). Examining infants' individuation of others by sociomoral disposition. *Frontiers in Psychology*, 10, 1271. <https://doi.org/10.3389/fpsyg.2019.01271>
- Thomson, J. J. (1985). The Trolley Problem. *Yale Law Journal*, 94(6), 1395–1415. <https://doi.org/10.2307/796133>
- Tinbergen, N. (1963). On the aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20(4), 410–433. <https://doi.org/10.1111/j.1439-0310.1963.tb01161.x>
- Tisak, M. (1995). Domains of social reasoning and beyond. In R. Vista (Ed.), *Annals of child development* (Vol. 11, pp. 95–130). Jessica Kingsley.
- Tomasello, M. (2018). The normative turn in early moral development. *Human Development*, 61(4–5), 248–263. <https://doi.org/10.1159/000492802>
- Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology*, 64(1), 231–255. <https://doi.org/10.1146/annurev-psych-113011-143812>

- Trivers, R. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46 (1), 35–57. <https://doi.org/10.1086/406755>
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
- Turiel, E. (1998). The development of morality. In N. Eisenberg (Ed.), *Social, emotional, and personality development. Volume 3 of the Handbook of child psychology* (5th ed., pp. 863–932). Wiley. Editor-in-Chief: W. Damon.
- Turiel, E. (2002). *The culture of morality: Social development, context, and conflict*. 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). Erlbaum.
- Turiel, E. (2018). Moral development in the early years: When and how. *Human Development*, 61(4–5), 297–308. <https://doi.org/10.1159/000492805>
- Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, 10(1), 72–81. <https://doi.org/10.1177/1745691614556679>
- Weisberg, M. (2005). Water is not H₂O. In D. Baird (Ed.), *Philosophy of chemistry: Synthesis of a new discipline* (pp. 337–345). Springer.