



Cognition in moral space: A minimal model

Bree Beal^{a,*}¹, Guram Gogia^{b,c,d}

^a Department of English, Clemson University, 815 Strode Tower, Clemson, SC 29631, United States

^b Department of Physics, Emory University, Atlanta, GA 30322, United States

^c Department of Environmental Systems Sciences, ETH Zürich, 8092 Zürich, Switzerland

^d Department of Environmental Microbiology, Eawag 8600 Dübendorf, Switzerland

ARTICLE INFO

Keywords:

Minimal model
Moral cognition
Moral psychology
Moral relationship
Moral space
Ontological framing

ABSTRACT

We describe moral cognition as a process occurring in a distinctive cognitive space, wherein moral relationships are defined along several morally relevant dimensions. After identifying candidate dimensions, we show how moral judgments can emerge in this space directly from object perception, without any appeal to moral rules or abstract values. Our reductive “minimal model” (Batterman & Rice, 2014) elaborates Beal’s (2020) claim that moral cognition is determined, at the most basic level, by “ontological frames” defining subjects, objects, and the proper relation between them. We expand this claim into a set of formal hypotheses that predict moral judgments based on how objects are “framed” in the relevant dimensions of “moral space.”

1. Introduction

Imagine an alien species on a distant planet X—let us call them Xlings—whose moral cognition is intimately tied to their perception of color. All objects are perceived by Xlings as either red, blue, or invisible, and color has striking moral implications for Xlings. They treat blue objects with careful consideration, and the bluer the object is, the stronger their sense of moral obligation toward it. In contrast, they are absolutely vicious toward red objects, and the redder an object is, the worse treatment it is felt to deserve. However, Xlings have no rigid rules about how to treat invisible objects—if an object is invisible, it may be used in whatever way anyone likes. Finally, this relation between color perception and moral judgment is complicated by an evident paradox, which we might call the paradox of “color realism.” Xlings implicitly understand color and its moral implications to be universal and consistent across time and context. But, as a matter of fact, color perception varies among Xlings and even within individuals over time. An Xling’s view of an object’s color may change gradually, following habits of mind and heart, but changes may also be rapid—A flash of anger might light up an object momentarily with red, or a warm feeling might imbue an object with blue. Yet, each time such a color transformation occurs, Xlings tend to implicitly assume that the new color they now perceive has been the object’s true color all along.

We can well imagine that the moral stakes of perceived color would lead to serious arguments among Xlings over the “true” color of various objects. Such arguments would be difficult to resolve, since interlocutors who disagreed about an object’s color would be speaking from experientially distinct realities. However, because of the flexibility of color perception, these differences would not necessarily be permanent: Various arguments, teachings, and practices would sometimes succeed in changing color perception. An Xling who had grown up seeing an object as red might begin seeing the object as less red. One might even change their own or another’s

* Corresponding author.

E-mail addresses: BreeLBeal@gmail.com (B. Beal), Guram.Gogia@emory.edu (G. Gogia).

¹ Permanent address: 1124 Virginia Ave NE, Atlanta, GA 30306, United States.

perception so that a formerly invisible object would become visible, or a red object might begin to appear blue—and vice versa.

Given the implications of changing color perception for moral cognition, Xlings might develop language and customs to reinforce or dismantle various color perceptions. They might coin slurs—creative ways of “painting things red”—and they might also develop terms of love and respect, meant to color objects with various shades of blue. Xlings would generally engage in such persuasive endeavors with the conviction of their implicit color realism—believing that the color they perceive at a given moment is the object’s true color, that the moral implications of the object’s color are universal, and that those who see it differently are definitely wrong.

Although human (“Earthling”) morality is not determined by color perception, we argue in this article that moral judgment on Earth is determined by a perceptual mode—“ontological framing” (see [Beal, 2020](#))—that is functionally analogous to Xling color perception. Ontological framing is just as determinative of Earthling moral judgment, and it is subject to an analogous paradox that might be called “framing realism”: Perceived framing and its moral implications are implicitly understood by Earthlings to be universal, timeless, and context-independent; however, object-framing varies among Earthlings and even within individuals over time.

Despite the essential role of ontological framing for determining moral judgment, Earthling moral psychologists have largely passed over framing in their attempts at modelling moral cognition. We seek to address this oversight by developing a “minimal model” ([Batterman & Rice, 2014](#)) that explains moral cognition in terms of the relation between ontological framing of “moral objects” and moral judgment. Our simplified model represents the essential dimensions of moral cognition as a kind of “moral space,” defined by the relevant “framing” dimensions. Below, we develop formal hypotheses for how an object’s properties, perceived along these dimensions, determine moral judgments.

2. Defining moral cognition

In Section 2, we define the primordial dimensions within which moral cognition emerges. We begin by identifying the conditions for a simple moral relationship between moral subjects and moral objects. We propose that perceived variance in moral objects along quantitative and qualitative dimensions gives rise to predictable variance in moral judgments. Once we have identified the morally relevant dimensions of object-perception, we turn in [Section 3](#) to the task of modelling the dynamics of moral cognition.

2.1. The primordial moral relationship

We follow [Beal \(under revision\)](#) in arguing that moral cognition is an emergent phenomenon, which derives from a combination of elements that, when taken in isolation, are nonmoral. This distinctively moral combination of elements we call the moral relationship: i.e., the relationship between that which is subject to moral norms (i.e., moral subjects) and that for the sake of which moral norms exist (i.e., moral objects).

We define the nonmoral conditions of moral subjects and moral objects below. However, first we caution that these 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 subjective “framings,” which may be imposed on any number of things.

This non-objective understanding of the moral relationship can help us appreciate the deeper structural similarity between moral judgments of behaviors, attitudes, and character traits. Before one can make a moral judgment, one must already implicitly know the answers to the following questions: “Who is subject to this moral judgment?” and “For what end?” In the case of a behavior norm, such as fair distribution, one may point to those who may be treated either fairly or unfairly—These are the “objects” of the moral relationship, answering the question “For what end?” In the case of one’s attitude, we may think of the behaviors that a bad attitude will tend to promote, such as disrespecting, neglecting, etc. various “objects.” In the case of moral character norms, there are attitudes and behaviors that moral and immoral character traits promote, which have consequences for various “moral objects.” One may even treat the possessor of a character trait as both a subject and object of the same moral norm, as in theories of virtue ethics (e.g., [Aristotle, 1934](#)) that propose that the subject who is morally obligated to cultivate virtue is also a moral object whose happiness and well-being depend on this cultivation.

With this phenomenological conception of the primordial moral relationship in mind, we now turn to the task of defining the dimensions within which moral subjects and objects emerge in cognition.

2.2. Moral objects and perceived inherent value (PIV)

Seeking to characterize the essential dimensions of a moral object, we approach moral phenomena from two different directions: asking, on the one hand, what is common to all moral philosophies that prescribe/proscribe behavior toward moral objects; and, on the other hand, what gives rise to the empirical symptoms of moral cognition that are studied by moral psychologists? This inductive method reveals one condition that is necessary, although not in itself sufficient, for distinguishing moral objects from nonmoral objects: perceived inherent value (PIV).

First, we look at moral metaphysics, or “metaethics.” It is important to appreciate the methodological nuance in how we interpret this literature. A metaethical theory is prescriptive, meaning that it proposes principles that help us figure out what we ought to do. There are diverse metaethical viewpoints, such as various forms of “consequentialism,” “deontology,” and “virtue ethics,” and they each point to somewhat different bases for a “valid” or “rational” moral metaphysics. By contrast, our descriptive model aims to describe how moral cognition works in everyone who possesses moral cognition—regardless of one’s metaethical commitments. Thus, for the purposes of our work, we bracket our own metaethical views. They are irrelevant. Yet, our metaethical agnosticism does not mean we ignore the field of moral philosophy—After all, the concept of ethics/morality is historically defined within philosophy—It

just means that we interpret this philosophical literature as moral psychologists. Rather than entering into prescriptive debates over the proper basis for a moral metaphysics—such as debates between consequentialists, deontologists, and virtue ethicists—we look for the features of moral cognition that are shared across all moral philosophies.

Every metaphysical theory of morality 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, 1934), “pleasure” (Bentham, 2010), “intrinsic value” (e.g., Moore, 1903; Korsgaard, 1983), “good-in-itself” (Kant, 1785/2002), “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. That is, they all invoke *some* idea of inherent value. Moreover, these moral philosophers all implicitly or explicitly distinguish their favored category of inherently valuable things from forms of noninherent value: e.g., “instrumental value” or “relative ends.” And these forms of noninherent value are supposed to be non-essential to morality. Thus, although philosophers disagree over which particular form of inherent value is the proper basis for a prescriptive moral metaphysics—Is it pleasure? A rational will? Eudaimonia?—they implicitly agree that one’s view of what is inherently valuable is essential for determining moral right from wrong. PIV is thus the name for the thing that all moral metaphysical theories share in common. Just as Xlings might disagree on the question of which shade of blue is most important while agreeing that *some* kind of blue is the basis for positive moral responsibilities, so do Earthling philosophers disagree on which type of inherent value is the proper basis for a moral system while implicitly agreeing that moral systems are determined by *some* kind of inherent value.

Next, we look at empirical evidence of moral cognition. PIV is also an implicit condition for the empirical criteria put forth by Social Domain Theory (SDT) (see reviews from Nucci, 2001; Smetana, 2006; Smetana, Jambon, & Ball, 2014; Tisak, 1995; Turiel, 1998, 2002; Turiel, 1998, 2002; 2006). Many studies show that children and adults of diverse cultures share a set of criteria for distinguishing moral rules from nonmoral conventions: Moral rules tend to be seen as i) universal and ii) context-independent and, less consistently, iii) serious and iv) punishable relative to social conventions—though there are gray areas (see, e.g., Heath, 2017). We note that, 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—since all other sources of normative obligation are in principle ruled out. As such, SDT criteria appear to imply PIV.

We acknowledge that the idea that SDT criteria indicate a distinctively moral cognitive domain is currently 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, Koller, & Dias, 1993; Kelly et al., 2007; Nichols, 2002; Nisan, 1987), concluding that moral judgment is probably not a distinctive area of cognitive performance. On its face, this argument is compelling. If morality cannot be defined reliably in terms of SDT criteria, and if we can find cases that are somewhere in-between moral and conventional domains (e.g., Heath, 2017), this would undermine an interpretation of SDT criteria as straightforwardly defining the moral domain.

While accepting this latter point, we propose that the key to better interpreting the evidence is to read SDT criteria not as a straightforward definition of the moral domain but rather as a set of “symptoms” of moral cognition. When making medical diagnoses, doctors appreciate that not all symptoms are equal. Similarly, when interpreting the evidence from SDT, moral psychologists should realize that not all the criteria have equal diagnostic value: Some suggest a specific condition, while others are more ambiguous. However, once we have determined the underlying etiology, such ambiguities can be explained.

Taking this diagnostic approach, we interpret the criteria of “seriousness” and “punishability,” as well as SDT’s claims about the value-content of moral judgments, as relatively ambiguous symptoms: Moral transgressions *tend* to be seen as more serious and punishable than violations of convention, and moral cognition *often* gives rise to concerns with justice, welfare, and individual rights—but there are exceptions.

Fortunately, there are two symptoms of moral cognition that are less ambiguous: Across many studies, perceived universality and context-independence distinguish morality from convention. As noted above, these two criteria together imply PIV—since, ruling out context and subjective opinion, the only remaining basis for moral obligation is the perceived value of the things themselves. It appears to us, therefore, that PIV underlies the various symptoms of morality that have been identified by SDT.

From this analysis, we conclude that PIV is a necessary condition for the emergence of moral objects. However, this does not mean that PIV is sufficient, in itself, to constitute a moral object. We explore this sufficiency problem in section 1.4 below, but first we turn to the task of defining moral subjects.

2.3. Moral subjects and perceived agency (PA)

By contrast to the inductive method used above to identify PIV, our next criterion can be identified through analytic deduction. When one is dealing with morality, one is dealing with normative standards concerning, for instance, how one ought to be or behave: One ought to do x rather than y , or one ought to be morally good, rather than evil. We observe that, before one can understand the normative obligation to “do x (rather than y),” one must already imagine that it is possible to do x —or, at least, that it is possible to try to do x . That is, one must perceive the one who is obligated to do x as possessing agency, in some minimal sense. Perceived agency (PA) is thus our second criterion of moral cognition.

In philosophical circles, agency is sometimes understood in a specialized sense of a “free will,” which implies that subjects are somehow independent of deterministic causal forces. However, in the present case, agency should 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 factors that make moral actions possible, such as personal abilities or situational exigencies.

One should bear this broad concept of agency in mind when interpreting 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, consequences influence moral judgments independently of an agent's perceived intentions (see [Martin & Cushman, 2016](#); [Patil et al., 2017](#)). Regardless of the (variable) extent to which cultures and individuals perceive moral subjects to be exercising a voluntaristic “free will,” it is nonetheless the case that any moral judgment depends upon the perception of some form of agency (PA)—some ability to be or behave as one ought.

We close this section with two caveats. First, we hope it is clear that we are not speaking about the objective or metaphysical reality of agency but, instead, about the subjective perception of agency. Second, we acknowledge that PA does not distinguish moral normativity from nonmoral normativity. For instance, nonmoral conventions may be expressed in exactly the same terms, which imply PA: e.g., “One should follow $\times$ [convention] rather than y [convention].” As such, although PA is necessary for moral cognition, this criterion does not distinguish morality from nonmoral normativity.

2.4. Moral objects are defined by “moral visibility”

If one accepts that PIV and PA are necessary conditions for moral cognition, this does not mean one must accept that PIV and PA are together *sufficient* for moral cognition. Let us consider the question of sufficiency by looking at each of our criteria in turn, starting with PIV.

Just because something is perceived as inherently valuable does not mean one considers themselves or others *morally* obligated toward that object. Indeed, we acknowledge that there are many cases where inherent value is perceived (PIV), and yet moral obligation is not acknowledged. The mere existence of such cases means that PIV is insufficient to define moral objects and, hence, moral cognition. There are two ways of addressing this insufficiency problem: the conventional way, which does not work, and a new way, which we propose below.

The conventional view, proposed first by moral philosophers and then adopted implicitly by moral psychologists, is that there must be a special category of PIV objects that is uniquely defining for morality (see, e.g., [Bentham, 1781/2010](#); [Kant, 1785/2002](#); [Korsgaard, 1983](#); [Scanlon, 1998](#)). In arguing for a categorical distinction between moral and nonmoral PIV objects, it is common to point to works of art and science, beautiful landscapes, and other impressive things as examples of objects that possess inherent value but that are not subject to moral consideration. We would expand the list, noting that nonmoral PIV could describe things both great and small—little bugs, for instance. Whatever the examples given, they have seemed to suggest that a broad conception of inherent value is insufficient to define moral objects. Accordingly, theorists have sought to identify the particular *kind* of perceived inherent value that gives rise to moral cognition (PIV_m), and to distinguish PIV_m from other kinds of PIV that do not give rise to moral cognition (PIV_{~m}).

Within the discipline of moral psychology, the search for a nonsubjective category of moral objects is doomed, however. Every objective-categorical PIV_m ever proposed is easily falsified by pointing to PIV objects outside the proposed category that some subset of moral subjects nonetheless view as moral objects. For instance, although [Kant \(1785/2002\)](#) famously defined moral objects in terms of the property of rationality (i.e., possessing a rational will), and there is a philosophical school that defends Kant's view to this day, it is hard to deny that at least some people at some times feel moral responsibilities toward something other than the rationality of rational beings. Another philosophical school, associated with [Bentham \(1781/2010\)](#), asserts that the ability to experience pleasure and pain is the defining feature of moral objects. Yet, once again, it is hard to deny that some people might at times feel moral obligations toward non-feeling objects. Thus, although it is perfectly acceptable for philosophers to defend either Kant's “deontological” or Bentham's “utilitarian” position in a metaethical argument over the *proper* basis for morality, it is nonsensical for a moral psychologist to do so when attempting to describe the *actual* basis for moral cognition wherever and whenever it happens to emerge. Whereas taking the Kantian position in a metaethical argument involves construing the utilitarian criterion for moral objects as somehow suboptimal, taking the Kantian position in order to define moral objects in cognition involves effectively denying that utilitarians exist. Because we acknowledge the existence of utilitarians and deontologists and other kinds of people, we abandon the hypothesis of an objective-categorical distinction between PIV_m and PIV_{~m}. That is, we reject the notion that there is any way to objectively distinguish between moral and nonmoral PIV objects—just as there is no way for Xlings to establish a category of objects that is visible for all Xlings.

Instead, we take a different approach, pursuing a nonobjective distinction between PIV_m and PIV_{~m}. Just as Xlings do not always agree about which objects are visible, Earthlings do not always agree about which objects are moral objects—Think, for instance, of controversies over the moral status of human fetuses, farm animals, rainforests, etc. Despite the obvious disagreements, it is still possible to identify the conditions under which an object achieves this status of “moral visibility” (i.e., being perceived as a moral object) in the mind of a given moral subject. Moral visibility might be determined by the *amount* of inherent value one attributes to the object and one's perceived *closeness* to the object, where “closeness” is understood not in a physical sense but in an existential-relational sense of intimacy/identity² (cf. Dasein's “spatiality” in [Heidegger, 1927/2010](#)). Just as objects are physically invisible if they are too small or too distant in space, so objects might remain morally invisible if their perceived value is too small or if they are perceived as existentially too distant from the subject (e.g., Is the fetus perceived as a full human or a part of the mother's body? Is the farm animal thought of as a sentient being or a material resource?). As such, PIV_{~m} would not be a categorically distinct species of PIV

² Although intimacy and identity are distinct representations, we posit that their moral function is the same. 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 similarity or identity with the object can create the necessary closeness (e.g., “they're human beings, just like me”).

object but, rather, a *status* of PIV objects whenever they fall below a subject's moral visibility threshold. By the same logic, "moral objects" would simply denote the status of PIV objects that cross the moral visibility threshold in the mind of a given subject.

The hypothesis that moral visibility is defined primarily by the perceived amount of inherent value and one's existential closeness to the inherently valuable object is just that—a hypothesis—albeit one that 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, Campbell, & Gaesser, 2021), which indicates that perceived closeness and/or perceived value shapes the amount and nature of our perceived moral responsibilities toward various objects. It should be possible to empirically investigate precisely what contributes to an object's moral visibility, and we propose ways of exploring this question below.

For now, however, we would like to make a more basic point: The proposition that moral cognition relies upon *some* conditional moral visibility is more than a mere conjecture. As long as the contrary hypothesis of an objective category of moral objects is false—and the above argument shows that it is false—the proposal that moral visibility is a subjective-conditional status must be true. We can investigate precisely what dimensions contribute to moral visibility, but such an investigation presupposes the truth of the generic "moral visibility" proposition.

Before we delve into ways of empirically investigating what contributes to moral visibility, consider a thought experiment that indicates why we think the proposed dimensions are plausible candidates. For some Earthlings, little bugs might be seen as possessing some small amount of inherent value, yet not warranting moral consideration. However, let us fiddle with our dimensions and see if we are able to bring a bug above the moral visibility threshold and thereby change its status from $PIV_{\sim m}$ to PIV_m .

Imagine a child who has the proposed attitude about little bugs—Bugs have some inherent value, but they are not worthy of moral consideration (i.e., for this child bugs are $PIV_{\sim m}$). However, we give our child a terrarium and a pet ant to observe in the terrarium. Supposing the child grows attached to the ant, thus increasing existential closeness, and begins to see their ant as possessing more inherent value than other little bugs—Is it not plausible that the child will then feel a sense of moral responsibility toward their pet ant? Will the child not feel guilty for squashing the ant, for instance? Similarly, suppose an Earthling who sees beautiful landscapes as $PIV_{\sim m}$ is fond of visiting a particular piece of wilderness that they find incredibly beautiful and toward which they develop a profound sense of intimacy—Is it not plausible that this person will begin to feel a moral responsibility to not desecrate this place? To respect, perhaps even to preserve and defend this beautiful landscape?

Such examples suggest to us that transforming the object along the proposed dimensions might be able change the object's status from $PIV_{\sim m}$ to PIV_m (ontological framing is just as malleable as Xling color perception) despite the absence of any objective-categorical distinction between $PIV_{\sim m}$ and PIV_m . Just as Xlings have no definitive categories of visible or invisible objects, Earthlings have no definitive categories of $PIV_{\sim m}$ and PIV_m objects. In both cases, the distinction is subjective, being determined by more basic criteria. In the case of Earthlings, we propose that moral visibility can be explained in terms of the amount of perceived inherent value and the subject's perceived existential closeness to the PIV object.

2.5. The "morality threshold" (MT) defines both moral objects and moral subjects

If PA and PIV_m are indeed necessary criteria for moral cognition, we must ask once again whether these criteria are sufficient to characterize the phenomenon of a simple moral judgment. Once again, however, it appears that our criteria are insufficient, as illustrated by the following hypothetical scenario.

Suppose Earthling Y's beloved child is swimming in the ocean when a shark attacks her. Is this action perceived as immoral? Let us say Y is ready to admit that the shark has exerted some form of agency in attacking their child. Specifically, it seems to Y that the shark attacked their child because it wanted or intended to, in some minimal sense—e.g., Y would distinguish this action from that of a remotely controlled robot shark, whose movements Y would attribute instead to the person with the remote. Hence, for Y, the condition PA is met in some minimal sense. In addition, it seems to Y that their child has a tremendous amount of inherent value and is a moral object. Thus, in Y's mind the criteria PA and PIV_m are met. Yet, for some reason, let us suppose that Y does not judge that the shark has done something immoral. Some unspecified condition, which is a necessary part of moral judgment, has not been met.

Just as it was necessary to nuance the criterion of PIV by introducing the conditional status of moral visibility, the above scenario shows that it is also necessary to nuance the PA criterion with a condition that we will call "moral vision." Although Y perceives the shark as an organism that acts intentionally, and in this sense exercises agency, Y does not expect the shark to see their child as a moral object (PIV_m). That is, Y does not think the shark possesses "moral vision"—at least, not with respect to Y's child. Although the shark possesses other perceptive faculties and can use them to act intentionally, Y does not think the shark can have awareness of the child as a moral object. This perceived absence of moral vision means that Y does not view the shark as being capable of exercising moral agency toward Y's child—i.e., Y sees the shark as possessing agency, but not *moral* agency ($PA_{\sim m}$) relative to the child. As such, although the event is certainly tragic, Y does not make a moral interpretation of the shark's action.

On the other hand, we can well imagine that if Y's neighbor were to attack their child, Y would register this behavior as a moral violation. The difference is that, in the latter case, Y does expect their neighbor to be capable of seeing their child as a moral object and understanding that attacking their child is immoral. The moral agency Y attributes to their neighbor (but not to the shark) emerges from the confluence of agency and moral vision that Y perceives their neighbor to possess. These conditions are together defining for perceived moral agency (PAm), or moral subjecthood.

Although we 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 is the moral visibility of an object 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; the subject must be perceived as possessing adequate moral vision; and the object must be perceived as being existentially close enough to the subject to warrant moral consideration. We propose that these factors conspire to define a threshold (MT) below which no moral relationship—and accordingly, no moral obligations—will be perceived. And we 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 of moral cognition.

2.6. PA, PIV, and MT are sufficient conditions for moral cognition

Finally, it seems to us that PIV, PA, and MT are together both necessary and sufficient conditions for moral cognition (see [Beal, under revision](#), for a more thoroughly developed version of this argument). If an object is perceived as deserving consideration for its own sake (PIVm), and a subject is perceived as being capable of striving to treat the object as it deserves (PA_m), then the formal synthesis of these perceptions is the inference that the subject ought to treat the object as it deserves to be treated. This perception of normative obligations between PA_m subjects (i.e., moral subjects) and PIV_m objects (i.e., moral objects) is itself the basic structure underlying all moral judgments.

Having identified these primordial dimensions of moral cognition, we now turn to the task of modelling the dynamics of moral cognition within a space that is defined by the proposed dimensions. In the following “minimal model,” we restrict our hypotheses to the relation between moral objects and moral judgments, while holding moral agency constant.

3. Modeling moral space

In Section 3, we elaborate the hypothesis that perceived moral responsibility toward “moral objects”—defined by PIV and MT—varies along several dimensions. Given our simplifying assumptions, we hypothesize that these dimensions determine both the amount of moral responsibility the subject feels toward the object and the perceived behavioral/attitudinal entailments of that responsibility. We emphasize that it is important to model these quantitative and qualitative dimensions of moral cognition separately, and we begin with the quantitative “amount” of perceived moral responsibility.

3.1. Amount of moral responsibility ($A_{MR}^{+/-}$)

We observe that there are two distinct kinds of perceived moral responsibility: positively valenced moral responsibilities “for” objects that are perceived to be inherently “good” (V^+ or “blue”) and negatively valenced moral responsibilities “against” objects that are perceived to be inherently “bad/evil/wicked” (V^- or “red”). We offer a single equation that expresses our initial hypothesis about how to approximate, under simplified conditions, the amount of each distinct kind of perceived moral responsibility ($A_{MR}^{+/-}$) within moral space.

Assuming a morally simple situation in which a generic “moral subject” (i.e., moral agency and relation to object are a constant) considers their responsibilities toward an isolated “moral object,” at a single moment in time, we propose that $A_{MR}^{+/-}$ is determined by the dimensions of Degree ($D^{+/-}$) (i.e., how much positive (D^+) or negative (D^-) inherent value one perceives a moral object to possess) and existential Relation ($R^{+/-}$) (i.e., how the moral subject relates to the moral object, where R^+ refers to existential “closeness,” defined by intimacy/identity, and R^- refers to existential “otherness,” defined by alienation/opposition). These hypothesized dimensions accord with both our intuition and mounting evidence (e.g., [Crimston et al., 2016; 2018](#); [Haslam, 2006](#); [Haslam & Loughnan, 2014](#); [Law et al., 2021](#)). Nevertheless, it is an empirical question whether Degree and Relation are, when taken together, truly the best predictors of direct measures of perceived moral responsibility ($A_{MR}^{+/-}$) between a subject and an object, and whether our intuition about the relation between these dimensions is approximately correct.

To formalize our prediction, let $D^{+/-}$ stand for the perceived amount of positive or negative inherent value possessed by the object and $R^{+/-}$ for the perceived existential closeness/otherness between the subject and moral object. We hypothesize that $D^{+/-}$ and $R^{+/-}$ have a positive relationship in moral space, and we express the quantity $A_{MR}^{+/-}$ as a function of $D^{+/-}$ and $R^{+/-}$ with a basic multiplicative formula: $A_{MR}^{+/-} = D^{+/-} * R^{+/-}$. This formula represents one way of predicting the actual amount of perceived moral responsibility between a subject and object ($A_{MR}^{+/-}$). When both D and R are positive, the formula yields a prediction for the perceived amount of positive moral responsibility “for” the object (A_{MR}^+). The same multiplicative formula also captures the relation between D and R when moral responsibility is negative. A negative D value represents the subject’s perception that the object is inherently bad (i.e., “red”). As such, the subject may feel negative moral responsibilities “against” the object, such as a responsibility to resist evil, punish transgressors, and so on. Our intuition is that perceived negative moral responsibility (A_{MR}^-) will increase as the “bad” object is perceived to be increasingly “other” (R^-), where “otherness” describes objects *against which* one defines oneself in opposition. As such, when both D and R are negative, the above equation yields a prediction of the amount of negative moral responsibility the subject feels *against* the “red” object (A_{MR}^-).

We anticipate that these dimensions will be highly correlated, such that few objects will be rated as positive on one dimension but

negative on the other. The above equation would yield a negative value in any such “incongruent” case. In reality, cases of incongruence are often highly fraught and ambivalent. For instance, when someone learns that a parent, sibling, child, or spouse has committed a terrible crime, they may have great difficulty deciding whether it is moral to protect their loved one or to act in the interest of a greater justice. Although we do not discuss such moral ambivalence in detail here, the phenomenon may offer anecdotal support for our hypothesis that Degree and Relation contribute independently to $A_{MR}^{+/-}$. In theory, a high negative score for $A_{MR}^{+/-}$ should predict a high degree of moral ambivalence.

Our equation for $A_{MR}^{+/-}$ may be tested against other formulas, to see which best predicts direct measures of a subject’s perceived amount of moral responsibility for or against an object. One might test hypotheses using the following procedure. Along the Degree vector, let D be represented by a number between -1 and 1 , where $D = -1$ expresses maximum negative value, $D = 0$ expresses the absence of negative/positive value, and $D = 1$ expresses maximum positive value. Similarly, let Relation be represented by a number between -1 and 1 , where $R = -1$ expresses maximum “otherness” of the object, $R = 0$ expresses the absence of otherness/closeness, and $R = 1$ expresses maximum existential closeness. D and R scores can then be obtained for experimental participants’ view of any number of objects. Using the hypothesized formula, $A_{MR}^{+/-}$ scores can then be derived for all objects. These scores can be compared to direct measures of participants’ felt or expressed moral responsibility toward the same objects. If a subject’s actual amount of positive or negative moral responsibility toward an object ($A_{MR}^{+/-}$) is well-approximated by a multiplicative relation between D and R , then an object’s $A_{MR}^{+/-}$ score should be highly predictive of $A_{MR}^{+/-}$. Specifically, assuming a sufficiently diverse object-set, $A_{MR}^{+/-} = D^{+/-} * R^{+/-}$ should be more predictive than either D or R alone (or other proposed dimensions).

It may be that an additive formula, such as $A_{MR}^{+/-} = a * D^{+/-} + b * R^{+/-}$, is a better predictor of real world moral judgments than is a multiplicative formula. However, we note that an additive formula would wash out incongruence and offer no explanation for the phenomenon of moral ambivalence. If the sign of $D^{+/-}$ and $R^{+/-}$ are opposite, the larger value will determine the sign of moral responsibility, and the amount of either positive or negative moral responsibility will be reduced. Thus, rather than representing incongruence as moral ambivalence, an additive formula would represent it as moral indifference.

Finally, whether a multiplicative or additive equation proves to be more precise, we may want to weight the dimensions differently—if it turns out that either $D^{+/-}$ or $R^{+/-}$ is a more important factor than the other. We propose an unweighted multiplicative equation for its simplicity and because it expresses the phenomenon of moral ambivalence. However, we look forward to empirical work that might nuance our proposed equation for $A_{MR}^{+/-}$.

Researchers may want to add one additional term to the hypothesized equation(s). As discussed above, it seems likely to us that subjects have a morality threshold (MT) expressing a function of D and R , which must be reached before they consider themselves or others to have any moral obligations at all toward a given object. As such, MT could be included in the formula as a constant calculated uniquely for each subject. Our hypothesized equation would then be expanded to $A_{MR}^{+/-} = D^{+/-} * R^{+/-} - MT$. It is an empirical question whether this additional variable adds predictive value that would outweigh parsimony considerations.

3.2. Qualitative entailments (E_i) of moral responsibility

Contemporary theories of moral cognition, such as Moral Foundations Theory (MFT) and the Theory of Dyadic Morality (TDM), propose that moral judgment involves a subject judging whether an action violates a principle of fairness, loyalty, authority, non-harm, etc. However, Beal (2020) points out that the meaning of these abstract principles is necessarily defined in terms of concrete “framing” of moral relationships. For instance, an action is not “disloyal” in abstraction, but only comes to fit this description under very specific conditions. For instance, it may be seen as “disloyal” for one to start dating a close friend’s recent ex-partner; yet, if one starts dating someone who has recently left a relationship with a total stranger, one will not feel they are being “disloyal” to the stranger. A stranger is, generally speaking, not the sort of object toward which loyalty is owed—though one may owe a stranger other forms of moral consideration, such as fair treatment, hospitality, etc.

Beal’s (2020) analysis suggests that moral cognition can be modelled without reference to abstract values like “loyalty.” If moral judgments necessarily involve the perception of concrete relationships, moral violations may be defined simply in terms of the (mis)fit between the moral expectations or entailments of a relationship and the action/attitude of a subject. As such, abstract values would seem secondary to the basic dynamics of moral cognition, nonessential side-effects of the more fundamental processes involved in everyday moral judgments.

The fundamental cognitive processes described in the present model involve both quantitative and qualitative inferences. In Section 3.1, we explore quantitative moral space, showing how subjects might gauge how much moral responsibility they owe to a given object. In this section (3.2), we explore qualitative moral space, showing how subjects determine which attitudes and behaviors are morally appropriate with respect to the object. In developing hypotheses about the qualitative dimensions of moral space, we engage in a deductive investigation of several abstract moral values. Our aim is to analyze the implicit logic of each value and to determine the kinds of concrete relationships that might give rise to each distinct moral quality.

We have already defined the valence (V) of moral responsibility, such that V^+ (i.e., a positive A_{MR}^+ score) indicates a positive moral responsibility “for” the (“blue”) object and V^- (i.e., a positive A_{MR}^- score) indicates a negative moral responsibility “against” the (“red”) object. Manifestly, there is a qualitative difference between positive moral responsibility and negative moral responsibility. In addition to V , we observe that there may be other morally relevant object-qualities. For instance, powerful deities might call for reverence and obedience, whereas vulnerable infants instead require guidance and protection. As such, even if one feels an equal amount of moral

responsibility toward two objects—such as a vulnerable infant and a powerful deity—the qualitative entailments (E_i) of that moral responsibility will vary with each moral object's perceived qualities.

Although we do not know the “true” number of relevant qualities, one may be able to identify candidate Quality dimensions with a novel method. By deductively exploring the varying moral entailments (E_i) of abstract values, it may be possible to infer some of the more fundamental object-qualities that shape these entailments. Identifying the morally relevant object-qualities is, in effect, identifying the relevant Quality dimensions of moral space. Toward this end, let us investigate the six abstract values described in Moral Foundations Theory (MFT) as “foundations” or “taste buds” of morality: Care/Non-harm, Justice, Loyalty, Respect for Authority, Sanctity/Purity, and Liberty (Haidt, 2012; Graham et al., 2013). We emphasize that the following exercise is not intended to be an exhaustive analysis but merely to illustrate our model.

Consider first the abstract value of Care. Care might seem a proper moral response toward a great many moral objects. Yet, what Care entails varies with the level of perceived vulnerability and power of the object. For instance, a vulnerable infant calls for a different form of care than an all-powerful deity. Thus, we suspect that Vulnerability and Power are relevant Quality dimensions, which shape a subject's sense of the morally proper attitudes and behaviors toward a given object.

Next, consider Justice. Justice is a very broad value that applies across much of moral space. Yet, what Justice entails varies depending on whether one is seeking justice for “good” (“blue”) objects or meting out justice against “evil” (“red”) objects. This observation reinforces the significance of object valence (V) for determining the appropriate moral attitudes and behaviors.

Manifestly, the entailments of Loyalty vary with the existential relation between a subject and object. Loyalty is something one owes to those who are close, but it does not apply to those we consider “other” (i.e., objects *against which* one identifies oneself in opposition). Thus, the value of Loyalty shows the qualitative significance of the Relation dimension of moral space. Because our model associates “otherness” with negative valence and “closeness” with positive valence, we may approximate this observation by reemphasizing the qualitative importance of V for determining morally appropriate attitudes/behaviors.

Next, consider Respect for Authority. One will feel a different quality of moral responsibility toward objects that are perceived as possessing legitimate Authority than toward similarly valuable and close objects that do not possess Authority: Consider, for instance, the different perceived entailments of the moral responsibility children feel toward their parents as opposed to that which they feel for their younger siblings. The basis for perceived legitimacy of an object's Authority might largely be captured by the subject's sense of the goodness (V^+) and power of the object. Yet, when we consider cultural norms of respect towards elderly people, even if they are frail and vulnerable, it strikes us that the perceived dignity of elders may have independent moral relevance. Nevertheless, for simplicity of illustration, we will not here posit a distinct “dignity” or “authority” dimension.

Next, consider Sanctity/Purity. The literature on Sanctity/Purity often points to sexual norms, which seem to have distinctive features. Although some sexual norms can be adequately explained in terms of perceived harm and imbalance of power—as in norms against sexual assault and statutory rape—other sexual norms evidently require a more nuanced explanation. For instance, sexual acts that seem objectively harmless—e.g., consensual, protected sex between siblings; casual pornography use; or even masturbation—may also be considered morally taboo. In such cases, some unique quality, such as the object's dignity may be at issue—whether that object is the object of a sexual act or whether the object is the subject themselves, as in taboos against masturbation (recall that a subject may be in a moral relationship with themselves). For simplicity, we do not include a distinct “dignity” or “purity” dimension in the present illustration.

Finally, consider Liberty. Some moral objects may be seen as inherently deserving of Liberty—e.g., inherently valued humans, nonhuman animals, and spirits—while for other objects the value of Liberty does not seem relevant. For instance, even if one feels a strong sense of moral responsibility to defend and cultivate a valued way of life, it does not make sense to argue that this “way of life” deserves Liberty. An evident difference between these object-classes is that people do not tend to perceive a “way of life” as being sentient, whereas people do attribute sentience to humans, nonhuman animals, and spirits. As such, it would seem that the value of Liberty is inherently *about* the proper moral response to objects that are perceived to be sentient.

With this crude method, we have deduced from the six proposed values of MFT three candidate dimensions of Quality—Vulnerability, Power, and Sentience—while also noting the qualitative significance of the perceived valence ($V^{+/-}$) of the moral relationship. We propose that the qualitative entailments (E_i) of moral responsibility between a subject and moral object will vary systematically as a function of how the subject perceives the object on (at least) these dimensions.

We express these hypothesized qualitative differences with archetypal descriptions, each of which represents the moral entailments (E_i) associated with a portion of moral space that is defined by V and a unique Quality dimension ($V^{+/-}$, Q_i). In future studies, researchers may employ empirical methods to establish the most relevant Quality dimensions (Q_i), along with their corresponding moral entailments (E_i), but for now we use our own preliminary hypotheses to illustrate how to use V and Q_i scores to locate the hypothesized E_i . Additionally, we suggest equations for determining the qualitative complexity (ρ) and combined amount and complexity (M), of perceived moral responsibility between a subject and moral object.

Set the maximum quantity for a given dimensional quality to $Q_i = 1$ and the minimum to $Q_i = 0$. This way, $Q_i = 0$ indicates that the object is perceived as lacking the quality in question, and $Q_i = 1$ indicates that the object is perceived as possessing a maximum amount of the quality. One can then correlate this Q_i score with empirically derived qualitative descriptions of the moral entailments (E_i) between a subject and object. For illustration, we propose hypothetical E_i descriptions to show what the Quality dimensions might entail.

Suppose that as Q_i approaches 1 on Vulnerability, objects that are perceived as entailing positive moral consideration (i.e., V^+) are felt to deserve protection/nurturance, whereas V^- objects are deemed worthy of punishment. These qualitatively different moral entailments—entailing protection/nurturance (E_1) and entailing punishment (E_2)—will then each be associated with an area within a

space defined by $V^{+/-}$ and Vulnerability. For illustration, we also hypothesize that as Q_i approaches 1 on Power, V^+ objects will be seen as deserving reverential respect and obedience (E_3), whereas V^- objects will be seen as deserving fearful hatred and resistance (E_4). Continuing, we hypothesize that as Q_i approaches 1 on Sentience, V^+ objects will be seen as deserving consideration of their feelings and desires (E_5), whereas V^- objects may be seen as deserving some sort of “mortification” (E_6), a concept we borrow from Catholic theology.

With a thorough questionnaire, researchers can empirically determine the relevant dimensions of Quality (the three above are very preliminary hypotheses) and obtain, for any population, descriptions of the various relevant entailments (E_i) of moral responsibility—e.g., protection, punishment, obedience, resistance, consideration, mortification, and so on. One can then locate these entailments within a space that is defined by the intersection of the $V^{+/-}$ quadrants with the vectors for all relevant Quality dimensions ($V^{+/-}$, Q_i). One might also set a quantitative threshold for Q_i scores (QT), which must be reached before a particular entailment (E_i) is considered effectively “in play.”

Once the relevant Quality dimensions have been established, we can use an individual subject’s Q_i scores to calculate the qualitative complexity (or “density,” ρ), and “mass” (M), of the moral relation between the subject and a given moral object. The equation for ρ emerges from the following considerations: Quality dimensions have a maximum Q_i score of 1. Thus, the maximum Q_i score for all Quality dimensions (Q_{max}) is equal to the number of Quality dimensions upon which an object is scored, which in the present case is 3. When more Quality dimensions are “in play” in a subject’s relation to a moral object, the moral relationship is qualitatively more complex, or “dense.” As such, qualitative density (ρ) can be expressed as a proportion of the sum of all Q_i scores for the object ($\sum_i Q_i$) relative to the total number of morally relevant Quality dimensions (Q_{max}): $\rho = \sum_i Q_i / Q_{max}$. Finally, we define the perceived moral “mass” (M) of the object with an equation that factors in both the quantitative amount of moral responsibility ($A_{MR}^{+/-}$) and the qualitative density of that moral responsibility (ρ): $M = \rho * A_{MR}^{+/-}$.

3.3. Hypothetical illustration of the model

By hypothesis, the perceived quantity and quality of moral responsibility between a moral subject and moral object may be approximated with the following equations:

1. Perceived amount of moral responsibility:

$$A_{MR}^{+/-} = D^{+/-} * R^{+/-} - MT$$

2. Perceived entailments of moral responsibility:

E_i above QT at ($V^{+/-}$, Q_i)

3. Perceived moral density:

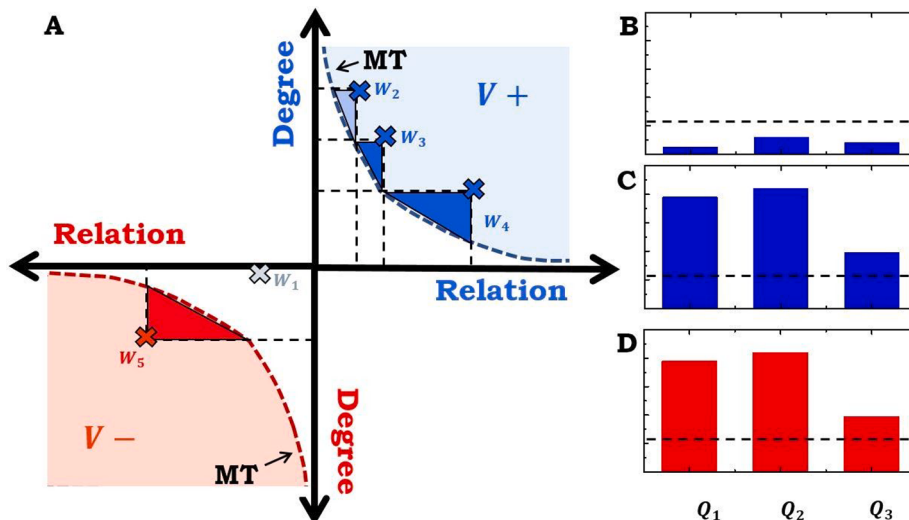


Fig. 1. A) Two primary axes of moral space, Degree and Relation, are represented on the y and x axis, respectively. The morality threshold (MT) is defined by an inverse relationship between Degree and Relation. The color blue represents positively valenced moral space (V^+), whereas red represents negatively valenced moral space (V^-) (white is outside moral space). The color saturation (inverse of transparency) of each triangle expresses the perceived “density,” (ρ) of the moral object as perceived by W_{1-5} (see Section 2.3). Q_1 , Q_2 , and Q_3 illustrate scores on the three Quality dimensions for each of W_{2-5} . B) corresponds to W_2 , C) corresponds to W_3 and W_4 , and D) corresponds to W_5 . *Some details of W_{1-5} in this figure differ slightly from the descriptions in Section 3.3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$\rho = \sum_i Q_i / Q_{max}$$

4. Perceived moral mass:

$$M = \rho * A_{MR}^{+/-}$$

Below, we show how one might use these equations to model a simplified moral relationship between an Earthling subject and moral object. First, we offer a caveat about the following illustration. A moral object can be any PIV object that the subject views as surpassing *MT*, including people (both individuals and groups), places/environments, animals, things, ideas, ways of life, and so on. Additionally, many moral objects are complex, being defined by multiple morally relevant “ontological frames” (see Section 4.2). As such, any moral subject is likely to live in a world populated by many moral objects of diverse kinds, and even one’s relationship to an isolated moral object may be complex. Finally, moral subjects may vary in the amount of agency and moral vision they perceive themselves or others to possess in any given circumstance, which may affect one’s level of perceived moral responsibility. However, we are bracketing for now the complexity that emerges from the large number of dynamic moral relationships one is likely to have, and instead restricting our analysis to “simple” moral relationships between generic moral subjects (i.e., relation to object and amount of perceived moral agency are held constant) and a single moral object in a single “frame” at a single moment in time.

Step into the minds of five white Americans in the early 19th century, as they consider their relationship to black people in America. Our first white American, W_1 , is an unabashed racist, who views black people as a categorically different race (i.e., low or negative on Relation) that is less than fully human (i.e., low on Degree). If our racist has a sufficiently high morality threshold (*MT*), Equation 1 will yield a negative value, indicating that black people are not worthy of moral consideration (see Fig. 1A). To our racist, black people will be morally invisible. W_1 might value black slaves for their economic worth (i.e., attributing instrumental value to them), yet, because W_1 neither values black people sufficiently for their own sake nor feels existentially close to them, W_1 does not view themselves as having any positive moral responsibilities toward black people (i.e., for W_1 , black people have the status of PIV_{-m}) (Fig. 2)

Next, consider a second white American, W_2 , who is not avowedly racist and instead professes that all people have equal value, but who nonetheless does accept the view, widely held by whites of the day, that whites and blacks belong to categorically different races. Assuming a modest *MT*, Equation 1 will yield a positive value for A_{MR}^+ , indicating that W_2 sees black people as worthy of moral consideration (i.e., for W_2 , black people are morally visible, having the status of PIV_m) (see Figs. 1A and 2). However, the R^+ value will be somewhat low, given the distance in racial identity, and this will bring down the A_{MR}^+ score, indicating that the amount of moral responsibility W_2 feels toward black people is moderate, rather than high. Because, in the mind of W_2 , black people exist in moral space, we can calculate the qualitative entailments (E_i), density (ρ), and mass (M) of the moral responsibility that W_2 feels toward black people in general. Suppose that W_2 rates black people as being low on Vulnerability, Sentience, and Power. Because no Q_i rise above *QT*, there are no hypothesized moral entailments, E_i (see Fig. 1B.). As such, the modest amount of moral responsibility W_2 feels toward black people is qualitatively empty and does not entail much in practice. Reflecting this qualitative emptiness, the quantities indicating the qualitative “density” (ρ) and “mass” (M) of W_2 ’s moral relation to black people, calculated from Equations 3 and 4, will be very low.

Next, consider a third white American, W_3 , who also does not consider themselves to be racist, but who also believes that whites and blacks belong to categorically different races. Assuming a modest *MT*, Equation 1 will yield a positive value for A_{MR}^+ , indicating that W_3 sees black people as worthy of moral consideration (i.e., for W_3 , black people have the status of PIV_m) (see Figs. 1A and 2). However, once again the R^+ value will be somewhat low, given the distance in racial identity, and this will bring down the A_{MR}^+ score, indicating that the amount of moral responsibility W_3 feels toward black people is moderate, rather than high. So far, W_3 has the same relation to black people as W_2 . However, suppose that W_3 rates black people as being high on Vulnerability and Sentience and moderate on Power. Given our hypothesized moral entailments, E_i (see Fig. 1C.), we infer that W_3 views black people as being worthy of a high amount of protection/nurturance and consideration of feelings/desires—bearing in mind that “high amount” is here a relative magnitude that remains within the quantitative limits on moral responsibility set by W_3 ’s moderate A_{MR}^+ . Additionally, we infer that W_3 feels a modest amount of reverential respect toward black people. The score indicating the qualitative “density” (ρ) of W_3 ’s moral relation to black people, calculated using Equation 3, will be high. However, Equation 4 will yield a moderate M , indicating modest overall “mass” of the moral relationship.

Now, consider a fourth white American, W_4 , who views black people as not only high in inherent value but also as belonging to the same human community as themselves. As such, Equation 1 will yield a high positive value for A_{MR}^+ , indicating that, for W_4 , black people are worthy of a high amount of moral consideration (see Figs. 1A and 2). Next, we can calculate the qualitative entailments,



Fig. 2. Ontological framing of black people by W_{1-5} , taken from Fig. 1. Size expresses $A_{MR}^{+/-}$ (bigger = higher $A_{MR}^{+/-}$). Color expresses V (blue = V^+ , red = V^-). Saturation expresses ρ (higher saturation = higher qualitative “density”). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

“density,” and “mass” of the moral responsibility that W_4 feels toward black people in general. Suppose that W_4 rates black people as being high on Vulnerability and Sentience and moderate on Power. Given hypothesized E_i (see Fig. 1C), we infer that W_4 views black people as being worthy of a high amount of protection/nurturance and consideration of feelings/desires, as well as a moderate amount of reverential respect. The quantity for all these positive entailments will be proportionally higher than similar entailments for W_3 , due to W_4 's higher A_{MR}^+ . As such, although Equation 3 will yield a high score for ρ that precisely matches W_3 's ρ score, Equation 4 will yield a high M , indicating that W_4 's relation to black people has a comparatively higher moral “mass” than W_3 's.

Finally, consider a fifth white American, W_5 , who views blacks as a subhuman and inherently “wicked” species. Supposing D^- and R^- have a high negative value, indicating that W_5 sees black people as highly “wicked” and extremely “other,” Equation 1 will yield a high value for A_{MR}^- , indicating that for W_5 black people are worthy of a high amount of moral consideration (see Figs. 1A and 2). However, we must not be deceived by the magnitude of the A_{MR}^- score but must pay attention to the valence (V) and qualities of moral responsibility felt by W_5 . V^- is negative (“red”), indicating that W_5 feels strong moral responsibilities *against* black people. Next, assume W_5 rates black people as being high on Vulnerability and Sentience, and moderate on Power. Given our hypothesized moral entailments, E_i (see Fig. 1D), we predict that W_5 will view black people as being highly worthy of punishment and mortification, and deserving a moderate amount of fearful hatred. Both ρ and M will be high, indicating that W_5 's negative moral relation to black people is high on both “density” and “mass.”

Now, suppose you are a black person living in early 19th century America. You meet a white person and have obvious reasons for wanting to know whether they are similar to W_1 , W_2 , W_3 , W_4 , or W_5 . Let us consider the information you could get from three theories of moral psychology: Moral Foundations Theory (MFT) (e.g., Graham et al., 2013; 2018; Haidt, 2012), the Theory of Dyadic Morality (TDM) (e.g., Gray, et al., 2012; Schein and Gray, 2015; Schein & Gray, 2015; 2018), and the present model.

A moral foundations questionnaire (MFQ) would indicate to what extent the white person holds the abstract values of Care, Justice, Loyalty, Authority, Sanctity, and Liberty. Given population differences in the extent to which people emphasize different values, you might be able to ascertain whether a given white person is politically conservative or progressive. However, you would not be able to learn fine-grained information about their sense of moral responsibility toward black people. For instance, with only information about the whites' values, you might be unable to discover any difference between W_{1-3} and W_5 .

The Theory of Dyadic Morality (TDM) explains immorality in terms of perceived harm committed by a moral “agent” against a moral “patient.” Since, in the examples above, we have not posited any behavior, whether harmful or harmless, it may seem that TDM can say nothing about W_{1-5} 's moral relation to black people. However, in TDM the relation between morality and harm can be read in multiple directions. In one direction, perception of harm implies perception of immorality, which implies something about the moral relationship between a moral “agent” and a moral “patient.” Since we have not posited any perceived harm, we can learn nothing by reading TDM in this direction from harm to immorality. However, in the opposite direction, one might be able to infer from the nature of the perceived moral relationship back to what sorts of behaviors will be judged as harmful and immoral. For instance, if we perceive someone as lacking “patience,” defined in terms of vulnerability and feeling (Schein & Gray, 2018), we might see them as less capable of being harmed, and this might have implications for our sense of what behaviors are (im)moral with respect to them. A description of this sort might be used to explain W_1 's attitude toward black people: W_1 might be said to “dehumanize” (see, e.g., Harris & Fiske, 2006; 2011; Haslam, 2006; Haslam & Loughnan, 2014; Leyens et al., 2003; Leyens, Demoulin, Vaes, Gaunt, & Paladino, 2007) or “super-humanize” (Waytz, Hoffman, & Trawalter, 2015; Wilson, Hugenberg, & Rule, 2017) black people such that they do not consider black people capable of being harmed. As such, W_1 might deny that the enslavement and exploitation of black people is harmful and, therefore, deny that such behavior is immoral.

The above explanation of W_1 's attitude toward black people accords more or less with the present model, except that TDM's explanation is relatively impoverished. Whereas our model describes a qualitatively diverse moral spectrum that includes a variety of both moral behaviors (e.g., protection, punishment, obedience, resistance, consideration, mortification, etc.) and immoral behaviors (e.g., neglect, cronyism, disobedience, capitulation, being inconsiderate, spoiling, etc.)—TDM attempts to capture this rainbow of qualitative complexity with only the vague concept of “harm.” As an explanation, the thesis that immorality = harm lacks nuance.

Beyond this qualitative impoverishment of explanation within TDM, numerous problems emerge from attempting to explain (im)morality in terms of harm. For instance, W_5 , who condones punishment and mortification of black people, would seem to embody a contradiction of TDM's central thesis. To square this circle, TDM's proponents might propose there has been a “dyadic reversal,” where W_5 does not perceive themselves to be an agent who is doing harm to black people, but instead considers themselves a vulnerable “patient” (Schein & Gray, 2018, p. 27). In order to maintain the thesis that perceived immorality = perceived harm, TDM must adopt a singularly flexible concept of “harm” such that moral subjects are said to deny the possibility of harming whatever is not perceived as a good moral object. For the same reason, TDM insists that perceived “agency” is contingent on moral judgment, such that people deny their own agency whenever they are fighting against perceived “evil” (Schein & Gray, 2018, p. 27): If one is not an agent, they cannot do anyone harm, and thus, they are not guilty of immorality.

Even if the above construal of what it means to be an “agent” and to perceive oneself as “harming” another is accurate, we might have difficulty using TDM to predict or explain all the differences in the moral relationships described above. For instance, TDM defines “agency” in terms of a certain form of Sentience (i.e., thinking) and Power, and “patience” in terms of a certain form of Sentience (i.e., feeling) and Vulnerability (see Schein & Gray, 2018). Since W_{3-5} all rate black people identically on Vulnerability, Sentience, and Power, it is plausible that W_{3-5} would also rate black people similarly on “agency” and “patience.” If so, one would not find resources within TDM to distinguish the (strikingly different) moral attitudes of W_{3-5} toward black people. Additionally, it is not clear how TDM would distinguish between the different forms of dehumanization of black people we observe in our hypothesized subjects: W_1 does not see black people as possessing a significant amount of inherent value; W_2 does not see black people as possessing

morally relevant qualities; and W_5 sees black people as possessing negative inherent value. Our model suggests that there are important moral differences in each respective subject's relation to black people, but a theory that measures moral cognition only in terms of perceived interpersonal harm would likely miss, or misconstrue, these differences.

By contrast to MFT and TDM, the present model offers detailed and specific hypotheses about the moral attitudes of $W_{1.5}$ toward black people. Three numbers express hypotheses for the amount and general nature of each white person's sense of moral responsibility toward black people: $A_{MR}^{+/-}$ quantifies the amount of moral consideration they feel is morally due to black people and indicates whether this moral attitude is that of someone who is for you (V^+) or against you (V^-). ρ quantifies the qualitative complexity, or "density," of the moral relationship as $W_{1.5}$ perceives it. And M combines the quantitative "amount" of moral responsibility and the qualitative "density" of the moral relationship to yield a measure of the overall moral "mass" of the moral relationship, as $W_{1.5}$ perceives it. In addition, archetypal descriptions (E_i) for all Q_i above QT indicate what each white person feels their moral relationship with you entails in practice. Thus, from these three numbers and the qualitative descriptions for all E_i "in play" at ($V^{+/-}$, Q_i), you gain a relatively detailed picture that might help you decide who to trust, how much to trust them, and in what circumstances to do so.

The distinctive feature of our model is that we take seriously the complex and changing views of individual subjects. By explaining moral judgments purely in terms of how objects are subjectively "framed" along a few relevant dimensions, the model is able to capture the radical differences in moral outlook, not only between but within groups, and not only within groups but within individuals over time. For instance, the present model can explain moral dynamism within individuals by showing how a change in framing may give rise to a change in the subject's sense of moral responsibility toward the object. We can imagine, for instance, that W_2 's view of black people could evolve over time to become closer to W_4 's; or, alternatively, that W_2 could become more like W_1 or even W_5 , if they adopted increasingly racist beliefs. We propose that such changes can and do occur even as one's abstract values or one's perception of harm remain roughly the same. In this subjective flexibility, we suggest, Earthling and Xling moral cognition are alike.

Remarkably, if our model is valid, one can glean all this information about the nature and amount of $W_{1.5}$'s sense of moral responsibility toward black people without making any inquiry into which abstract moral values they favor, without asking them to solve any moral dilemmas, and (aside from perhaps establishing their baseline MT) without asking them to make any moral judgments. Of course, you might gain similar information by asking $W_{1.5}$ directly about their moral responsibilities toward black people. However, we propose to *predict* moral judgments based on "framing" information (analogous to Xling color perception) and, in doing so, to explain the deeper cognitive dynamics that give rise to these judgments. If our equations (or related formulas using the proposed dimensions of moral space) predict actual moral attitudes and behaviors across diverse contexts and relationships, this evidence will support our theory that Earthling moral cognition is fundamentally determined by "ontological framing."

4. Next steps

We conclude the article by proposing some procedures for testing our hypotheses and, accordingly, refining or abandoning the present model (see [Tables 1 and 2](#)). Finally, we consider next steps for connecting this simplified model to more complex analyses.

4.1. Testing the model

[Table 1](#) offers suggestions for how to operationalize the dimensions of moral space. An overall $D^{+/-}$ score for an object can be calculated by subtracting the D^- score from the D^+ score. Similarly, an overall $R^{+/-}$ score can be calculated by subtracting the R^- score from the R^+ score. Alternatively, to analyze something as a "complex" object that inhabits multiple frames, treat D^- and R^- separately from D^+ and R^+ , and calculate two distinct $A_{MR}^{+/-}$ scores for the same object. Such a procedure might offer a different way modelling moral ambivalence, wherein a moral subject feels both positive and negative moral responsibilities toward the same object.

One caveat to the suggestions in [Table 1](#) is that nuanced distinctions, such as that between inherent value and non-inherent value, may be difficult for participants to understand. If so, experimentalists will need to find ways of adequately explaining or intuitively operationalizing these concepts.

[Table 2](#) offers several ways of comparing the predictions of our model to "direct measures" of perceived moral responsibility. If our model is approximately correct, and if the "direct" measures of moral cognition are approximately accurate, our equations should be highly predictive of direct measures.

Table 1
Operationalizing D , R , and Q_i

Dimensions	Operationalizations
D and Q_i	How much of each quality does O possess? 1. Inherent goodness 2. Inherent badness/wickedness/evil 3. Power 4. Vulnerability 5. Sentience. 6. etc.
R^+	How close are you and O, in the sense of intimacy or identity? OR
R^-	To what extent would you feel personally offended if you heard someone saying derogatory things about O? To what extent do you think of O as something/someone against which you define yourself in opposition? OR
	To what extent would you feel insulted to learn that you are associated with O?

Table 2
Validating the Model.

	$A_{MR}^{+/-}$ predicts $A_{MR}^{+/-}$	E_i predicts E_i
Selection/ Ranking	Ask participants to rank objects by the amount of moral responsibility ($A_{MR}^{+/-}$) owed to each. $A_{MR}^{+/-}$ scores, derived from Equation 1, should predict within-subject scores for the same objects, and variation in $A_{MR}^{+/-}$ scores for each object across participants should predict variation in $A_{MR}^{+/-}$ scores for the same objects.	Ask participants to choose qualitatively appropriate moral attitudes or behaviors (E_i) toward objects that have matched within-subject $A_{MR}^{+/-}$ scores. (V_i, Q_i) scores correlated with hypothesized E_i should predict E_i for the objects within-subject, and variation in hypothesized E_i across subjects should predict variation in E_i when $A_{MR}^{+/-}$ scores are matched.
Moral Dilemmas	Ask participants to decide between competing moral obligations toward objects that they have previously judged as qualitatively similar (i.e., similar (V_i, Q_i)). Because objects are matched on quality, participant choices may be interpreted as an expression of relative $A_{MR}^{+/-}$. As such, $A_{MR}^{+/-}$ scores should predict which objects participants choose to help, obey, respect, etc. (researchers must choose an appropriate E_i).	Ask participants to decide between competing moral entailments (E_i) toward objects that have similar within-subject $A_{MR}^{+/-}$ scores. By hypothesis, E_i will be predicted by E_i scores (as determined by (V_i, Q_i)) within-subject, and variation in E_i across subjects will be predicted by variation in E_i when $A_{MR}^{+/-}$ scores are matched.
Behavioral Measures	Measure differences in the readiness to help or the amount of help given to objects that are matched within-subject on E_i . If E_i is controlled, participant choices may be interpreted as an expression of relative $A_{MR}^{+/-}$, and these choices should therefore be predicted by $A_{MR}^{+/-}$.	Observe differences in quality of moral response (E_i) toward objects that are matched within-subject on $A_{MR}^{+/-}$. Behavioral variation in E_i should be predicted by E_i .

Another approach to validating the model would be to test for a causal relationship by experimentally transforming subjects' perception of objects along $D^{+/-}$ and/or $R^{+/-}$ (i.e., altering objects' perceived "color"). With significant changes along these dimensions, one may even be able to move objects into or out of moral space (i.e., making objects morally "visible" or "invisible"). Such experiments could be used to test whether the proposed dimensions are in fact defining for moral objects, and also to (in)validate the *MT* hypothesis. If it is possible to significantly transform PIV objects along $D^{+/-}$ and $R^{+/-}$ without affecting their moral visibility and without altering $A_{MR}^{+/-}$, this would disconfirm our model. If, on the other hand, we observe a tight linkage between $A_{MR}^{+/-}$ and $A_{MR}^{+/-}$, such that changes in the former lead to corresponding changes in the latter, these findings would support our model.

We emphasize that the proposed approaches to testing or validating our model are preliminary suggestions. We trust that experienced experimentalists may develop more sophisticated methods than we can currently envision.

4.2. Complexifying factors

The present model attempts to illustrate how moral cognition works at the most basic level, in a "simple" relation between a single moral subject and a single moral object. However, moral cognition in everyday life may be complicated by additional factors, including the following:

1. Complexity of moral relationships. Moral objects are defined in relation to moral subjects, and some forms of relation are difficult to characterize with the present approach that emphasizes object qualities. Consider, for instance, perceived ownership. Moral obligation places limits on the pursuit of one's desires or interests. But ownership may be perceived as a kind of limit on such limitations: If one claims that something is their property, they may be implicitly rejecting limits on how they may treat their "property." Think of the way some people feel about their own bodies: If someone feels that their body is entirely their own property, they may also feel that no one has a right to place moral limitations on how they treat their body. This rejection of moral limitation has nothing to do with the body being a special kind of object. After all, the same person may have rigid principles regarding how *others* may treat their body. Instead, this limitation on limitations may simply be a feature of the perceived "ownership" relation.

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* 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. Additionally, 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.

A more complex formula for $A_{MR}^{+/-}$ could account for this inverse relationship between perceived ownership and perceived moral responsibility by adding a "perceived ownership" (O) variable into the equation: e.g., $A_{MR}^{+/-} = \frac{D^{+/-} \times R^{+/-}}{O} - MT$, where O is a number greater than or equal to 1. In the above equation, $O = 1$ represents no perceived ownership, and it thus yields back the original equation. However, higher values for O lead to the eventual disappearance of perceived moral obligation, as $A_{MR}^{+/-}$ falls below zero.

Although it may be possible to account for perceived ownership using the above approach, there are other kinds of relationship that present unique difficulties. Norms defining appropriate sexual objects and behaviors are like other moral norms in that they emerge from perceived (mis)fit between the moral entailments of a relationship and the nature of one's actions. Yet, it would be difficult to define moral norms in terms of a few object properties. What are the properties of an appropriate sexual object? The answer to this question varies greatly depending on who the subject is. What is an appropriate sexual behavior? One must know the relevant context.

The present “minimal” approach assumes a generic moral subject and defines (im)moral behavior in terms of the perceived properties of moral objects. The example of sexual norms reveals the limitations of this approach—not because the “ontological framing” thesis is incorrect but because framing is fundamentally relational, and the assumption of a generic “moral subject” limits our analysis. We look forward to future work investigating the complex interplay between subject and object framing in moral relationships.

2. Diversity of moral subjects. People are differently constituted and differently abled, and judgments about morally appropriate behavior vary with subjects’ perceived qualities/abilities. To account for this complexity of moral subjects, one may introduce the amount of perceived moral agency (PA_m) to the equation for $A_{MR}^{+/-}$, yielding a nuanced equation: $A_{MR}^{+/-} \rightarrow A_{MR}^{+/-} * PA_m$, where $0 \leq PA_m \leq 1$. In this equation, $PA_m = 0$ represents no perceived moral agency, and $PA_m = 1$ represents maximal perceived moral agency.

3a. Complexity of moral objects. A single object may occupy many moral frames. For instance, one may be a wife, mother, friend, enemy, authority figure, subordinate, and so on. Thus, a full model of moral cognition must account for the complexity of a subject’s moral relationships to objects that occupy multiple ontological frames, many of which may be subject to dynamic change. One’s obligations to the object-as-X might conflict with one’s obligations to the object-as-Y. For instance, when someone hires their child to work for them, a moral responsibility to care for the child may conflict with a moral responsibility to treat all employees fairly. This problem of nepotism is in part a conflict between two relationships—the parent relationship and the employer relationship—toward the “same” complex object (child/employee).

3b. Complexity/Interaction among moral objects. Subjects have moral relationships with multiple objects, and we may experience zero-sum competition between them, such that our moral responsibility toward one object diminishes our sense of responsibility toward another object. In the nepotism example from 3a, the parent/employer must not only balance their responsibilities toward their complex child/employee object, but must also consider competition between their child/employee and other employees whenever the company goes through layoffs, decides who to promote, etc. Although the current simplified model does not consider inter-object moral complexity, these ecological dynamics of moral space are worthy of analysis and investigation.

4. Finally, the several factors described above are all forms of ontological framing that directly influence moral cognition. A full model of moral psychology would also show how moral cognition is affected by indirect variables, such as social, environmental, and evolutionary factors. In doing so, it is crucial to appreciate that the influence of these factors on moral cognition is mediated by ontological framing. For instance, environmental exigencies may transiently or permanently change the object-frame in morally relevant ways. An object that is perceived as powerful may be vulnerable in a particular situation, or may become vulnerable due to changing social factors, sudden physical disability, etc. These situational exigencies affect one’s sense of what is morally appropriate precisely because they influence morally relevant features of the object-frame.

5. Conclusion

Throughout this article, we have developed an analogy between Earthling and Xling cognition—proposing that, in the same way that researchers studying Xling moral cognition might start by mapping the relation between color perception and moral judgment, researchers studying Earthling moral cognition should start by mapping the relation between ontological framing and moral judgment. We have offered just such a map, represented as a “moral space” that contains both quantitative and qualitative dimensions. We conclude by considering how our model relates to current moral psychology research. What follows is not an exhaustive literature review, but a brief survey of what we take to be the most salient approaches to understanding moral cognition in the current field. Our intent is not to criticize existing theory, but to show the crucial sense in which the current model departs from all that has come before.

The vast literature on moral dilemmas (e.g., Foot, 1967; Greene, 2013, 2015; Kohlberg, 1981; Mikhail, 2007; Thomson, 1985) asks participants to choose between conflicting responsibilities toward multiple inherently valuable or, as it were, “blue” objects. But this literature overlooks the question of how an object comes to be perceived as a “blue” in the first place. In a similar vein, research into “dehumanization” (e.g., Harris & Fiske, 2006; 2011; Haslam, 2006; Haslam & Loughnan, 2014; Leyens et al., 2003, 2007) tacitly assumes the “blueness” of a certain class of objects (i.e., humans), while overlooking morally relevant distinctions, such as the difference between being “invisible” (i.e., not being perceived as a moral object) and being “red” (i.e., being perceived as a “bad” moral object). To its credit, the dehumanization literature exposes some of the flexibility of Earthling moral perception. But the incomplete attention paid to other morally relevant framings—beyond that of being “human”—leaves the relevant theory underdeveloped. One cannot simply extrapolate from research on dehumanization to a model such as ours.

We believe that a further step in the right direction is taken by researchers looking at moral expansiveness (e.g., Schwartz, 2007), or the inclusiveness of Earthling moral circles (e.g., Crimston et al., 2016; 2018). This research shows how an Earthling’s existential closeness to an object influences our sense of moral responsibility toward that object. Yet, this literature does not rigorously distinguish between subjective and objective kinds of closeness, and therefore does not convey our claim that only subjective closeness is morally decisive. For instance, what is morally relevant is not whether or not one is biologically related to another. After all, it is possible to be entirely unaware of one’s relatives, and one cannot feel moral obligations toward something one knows nothing about. Conversely, it is possible to feel that someone biologically unrelated, or even a member of another species, is one’s family. It is the subjective perception of closeness that decides moral cognition, and objective delineations of closeness are only relevant to the extent that they predict subjective perceptions. We add that the literature on moral circles does not identify and incorporate the other morally relevant features of perception (PA_m , PIV_m, and MT) that we describe above. Thus, while this important research is consistent with the present analysis, it does not offer the distinctive features of our model.

Finally, as shown in Section 2.3 above, the two (arguably) most influential theories of moral cognition at present—MFT and

TDM—both fail to adequately consider the relation between ontological framing and moral judgment. MFT skirts the issue entirely, while TDM inappropriately attempts to reduce the several morally relevant dimensions of ontological framing to a single dimension of perceived interpersonal harm.

In fact, the essential features of the present model are new. We draw upon and sometimes reinterpret the existing literature, but within this literature, no one has proposed that moral cognition can be reduced to dynamics of ontological framing, nor posited what these dynamics might be. Recently, a few theorists (Rottman et al., 2020; Schein, 2020) have noted a parallel between the idea of ontological framing and Turiel et al.'s (1987) concept of “informational assumptions.” We acknowledge that “informational assumptions,” or simply beliefs, are a big part of ontological framing. But one's beliefs are not the whole story. Subjects with identical beliefs may feel different moral obligations toward the same object, because they inhabit different ontological frames. For instance, two individuals (A & B) with identical beliefs will likely perceive different sexual norms toward a third person (C), if C happens to be A's sibling. The difference in A and B's sense of what is morally appropriate with respect to C does not come from a divergence in “informational assumptions,” but from the fact that they each have a different moral relationship to C (i.e., being a sibling vs. not being a sibling). This specificity of one's relationships is decisive for morality, independent not only of one's abstract values but also of one's “informational assumptions” about the world.

The present model emerges from a recognition that moral cognition is defined by the perception of inherent value (PIV), and that the quantity and quality of perceived moral responsibility is determined by ontological framing of PIV objects in a uniquely moral cognitive space. We reject as untenable the view that there is, or must be, an objective-categorical distinction between moral and nonmoral objects. Instead, we propose a morality threshold (*MT*) that PIV objects must reach before a subject will view them as (“morally visible”) moral objects. Once the object has crossed *MT*, it is effectively in moral space and subject to moral consideration: i.e., it has the subjectively defined *status* of PIV_m. The equations proposed in this minimal model approximate our initial intuitions about how ontological framing molds moral responsibility in the several dimensions of moral space. We anticipate and welcome empirical insights or theoretical considerations that would further refine these hypotheses.

CRedit authorship contribution statement

Bree Beal: Conceptualization, Writing - original draft. **Guram Gogia:** Writing - review & editing, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We warmly thank Adam Bear, Elspeth Brotherton, Ben Hill, Heather Saul, and two anonymous reviewers for thoughtful comments on early drafts of our manuscript.

References

- Aristotle. The Nicomachean ethics. (1934). E. Capps, T.E. Page, & W.H.D. Rouse (Eds.) (H. Rackham, Trans.). London: William Heineman.
- Barrett, H., Bolyanatz, A., Crittenden, A., Fessler, D., Fitzpatrick, S., Gurven, M., ... 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>.
- Batterman, R., & Rice, C. (2014). Minimal model explanations. *Philosophy of Science*, 81(3), 349–376. <https://doi.org/10.1086/676677>.
- 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. Under revision. The nonmoral conditions of moral cognition. *Philosophical Psychology*.
- Bentham, J. (2010). An introduction to the principles of morals and legislation. White Dog Publishing (originally published in 1781).
- 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. R., Hornsey, M. J., Bain, P. G., & Bastian, B. (2018). Toward a psychology of moral expansiveness. *Current Directions in Psychological Science*, 27(1), 14–19. <http://doi.org/10.1177/0963721417730888>.
- 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. and Laurence, S. (2015). Moral parochialism and contextual contingency across seven disparate societies. *Proceedings of the Royal Society B*, 282(1813) (August 22). doi: 10.1098/rspb.2015.0907.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5–15.
- 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. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>.
- Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwicz, 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.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23(2), 101–124.
- Greene, J. (2013). *Moral tribes: Emotion, Reason and the gap between us and them*. New York, NY: Penguin Press.
- Greene, J. D. (2015). The rise of moral cognition. *Cognition*, 135, 39–42.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon Books.
- 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. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, 17(10), 847–853.

- Harris, L. T., & Fiske, S. T. (2011). Dehumanized perception: A psychological means to facilitate atrocities, torture, and genocide? *Zeitschrift für Psychologie*, 219(3), 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. <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.). Albany: State University of New York Press (Original work published 1927).
- 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(2), 117–131. <https://doi.org/10.1111/j.1468-0017.2007.00302.x>.
- Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of justice* (Vol. 1). New York, NY: Harper & Row.
- Korsgaard, C. M. (1983). Two distinctions in goodness. *Philosophical Review*, 92(2), 169. <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>.
- Leyens, J.-P., Cortes, B., Demoulin, S., Dovidio, J. F., Fiske, S. T., Gaunt, R., ... Vaes, J. (2003). Emotional prejudice, essentialism, and nationalism: The 2002 Tajfel lecture. *European Journal of Social Psychology*, 33(6), 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.
- Martin, J. W., & Cushman, F. (2016). Why we forgive what can't be controlled. *Cognition*, 147, 133–143.
- 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>.
- Moore, G. E. (1903). *Principia ethica*. Cambridge, England: Cambridge University Press.
- Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, 84(2), 221–236.
- Nisan, M. (1987). Moral norms and social conventions: A cross-cultural comparison. *Developmental Psychology*, 23(5), 719–725.
- Nucci, L. P. (Ed.). (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, 5200.
- Plato. 'Protagoras' and 'Meno'. (2004). (R. Bartlett, Trans.). Ithaca: Cornell University Press.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Belknap Press.
- Rottman, J., Zizik, V., Minard, K., Young, L., Blake, P. R., & Kelemen, D. (2020). The moral, or the story? Changing children's distributive justice preferences through social communication. *Cognition*, 205, 104441.
- Scanlon, T. (1998). *What we owe to each other*. Cambridge, MA: Belknap Press.
- Schein, C. (2020). The importance of context in moral judgments. *Perspectives on Psychological Science*, 15(2), 207–215. <https://doi.org/10.1177/1745691620904083>.
- Schein, Chelsea, & Gray, Kurt (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, 32–70. <https://doi.org/10.1177/10888683176982>.
- Schwartz, Shalom H. (2007). Universalism values and the inclusiveness of our moral universe. *Journal of Cross-Cultural Psychology*, 38(6), 711–728.
- 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.
- Stich, S. (2018). The quest for the boundaries of morality. In K. Jones, M. Timmons, & A. Zimmerman (Eds.), *The Routledge Handbook of Moral Epistemology*. New York: Routledge.
- Thomson, Judith Jarvis (1985). The Trolley Problem. *Yale Law Journal*, 94(6), 1395. <https://doi.org/10.2307/796133>.
- 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.
- 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.
- Waytz, Adam, Hoffman, Kelly Marie, & Trawalter, Sophie (2015). A superhumanization bias in Whites' perceptions of Blacks. *Social Psychological & Personality Science*, 6(3), 352–359.
- Wilson, John Paul, Hugenberg, Kurt, & Rule, Nicholas O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology*, 113(1), 59–80.