

# AGAINST AND FOR ETHICAL NATURALISM

## Or: How Not To “Naturalize” Ethics

Berit Brogaard and Michael Slote

### ABSTRACT

Moral realism and ethical naturalism are both highly attractive ethical positions but historically they have often been thought to be irreconcilable. Since the late 1980s defenders of Cornell Realism have argued that the two positions can consistently be combined. They make three constitutive claims: (i) Moral properties are natural kind properties that (ii) are identical to (or supervene) on descriptive functional properties, which (iii) causally regulate our use of moral terms. We offer new arguments against the feasibility of Cornell realism and then show that there is a way to be a naturalistic realist that avoids internal inconsistency and uniquely provides for moral normativity.

### I. INTRODUCTION

The last three decades of the twentieth century witnessed a sudden revival of the desire to “naturalize” ethics, epistemology, and other areas of value theory alongside notoriously elusive concepts like *intentionality* and *reference* (Donnellan 1966; Quine 1969; Kripke 1970/1980; Kim 1988; Kitcher 1992; Putnam 1995; Almeder 1990; Antony 1993; Dretske 1995; Tye 1995). One of these revival efforts was particularly ambitious. It claimed to give us what everyone else seemed to think one cannot have: moral realism and ethical naturalism. This metaethical position quickly acquired the nickname (or “term of endearment,” if you will) “Cornell realism,” because most of its main advocates at the time, viz. Richard Boyd (1988), David Brink (1989; 2001; 2007), and Nicholas Sturgeon (1985) studied or worked at Cornell University.<sup>1</sup>

To get a better sense of what the Cornell realists were up to, it will be instructive to dwell for a moment on how to understand

moral realism and ethical naturalism, independently of one another. Let’s start with moral realism. To a first approximation, moral realism is the view that moral judgments are akin to judgments of facts, like “1989 marks both the ratification of the U.S. Constitution and the beginning of The French Revolution,” “Most bank robberies take place on Fridays,” or “Iguanas have three eyes, one of which is located on top of their head,” in being objectively true or objectively false (here: they are all objectively true). To put it differently: like factual judgments, moral judgments have a truth-value independently of people’s moral attitudes and dispositions and independently of cultural rituals and social conventions.<sup>2</sup> One of the attractions of this metaethical stance is that it coheres with our tendency to engage in sincere and vigorous moral debate rather than respond with indifference to people with moral convictions that are at odds with our own. If moral judgments were, say, mere expressions of personal taste or bodily

sensation, and we were fully aware that this were so, then a statement like “Adding alluring flavors to e-cigarettes is wrong” should be no less likely to stir up heated emotions or ferocious debate than expressions of personal taste or bodily sensations such as “I don’t like broccoli,” “surströmming makes me nauseous,” “that woolen sweater makes me feel itchy” or “fried food gives me heartburn.”

Ethical naturalism, the other strand of Cornell realism, is the view that moral properties of people, actions, or institutions, such as “right,” “wrong,” “good,” and “bad,” are neither supernatural properties, such as the property of being condoned by God, nor *sui generis* evaluative properties, as argued by G. E. Moore (1903) and Phillipa Foot (2001). Rather, they are natural kind properties much like garden variety natural kind properties such as *being Gold*, *having mass*, or *having a headache*, except that they are likely to be a lot more complex than other natural kind properties. One of the chief virtues of ethical naturalism is that it takes the mystery out of ethics. Given naturalism, ethical inquiry is akin to other forms of scientific inquiry in asking which properties are best suited for the role of explaining and predicting observations—in this case observations of action.

Cornell realists frame their position as a rejection of the eliminativist view that moral language in principle could be replaced by purely descriptive (or “non-moral” or “natural”) language without any explanatory loss (see e.g., Boyd 1988, p. 194). In other words, for any moral statement, there is a purely descriptive statement with the same explanatory power, such that the latter in principle could be logically deduced from the former. Both Moore and the Cornell realists reject moral eliminativism. To borrow a term of philosophy of mind, we can say that Moore and the advocates of Cornell realism are equally committed to an explanatory gap between moral and descriptive judgments regarding human action.<sup>3</sup>

But Cornell realists pride themselves on being ethical naturalists. So, their rejection of moral eliminativism is not what sets them apart from Moore. Nor is it their take on the supervenience of (or identity to) moral properties on descriptive properties, that is, there cannot be a difference in moral properties without a difference in descriptive properties (McPherson 2019). (Here we assume an abundant conception of *moral properties* according to which they are the semantic values of moral predicates, that is, the contribution the predicate makes to the proposition expressed by the sentence in which it occurs—this is in line with Moore’s use of “concept” as synonymous with “property”).

Granted, members of the Cornell school hold that moral properties morally supervene on descriptive functional properties insofar as the former are identical to the latter, whereas Moore held that moral properties are *sui generis* and therefore are not identical to descriptive properties, but as we will see, taking moral properties to be identical to descriptive functional properties is not essential to this brand of naturalistic moral realism.<sup>4</sup>

What sets Cornell realism apart from Moorean non-naturalism is that they extend the causal theory of reference advocated by Saul Kripke (1970) and Hilary Putnam (1975) to moral terms, and then explain why moral properties morally supervene on (or are identical to) descriptive properties.<sup>5</sup> On this view, moral terms like “right” and “wrong,” refer to the properties initially dubbed “right” or “wrong,” as long as the term has been passed on in an uninterrupted fashion from speaker to speaker, each of whom has intended to refer to whatever the previous speaker was referring to.

Extending the causal theory of reference to moral terms has proven difficult. Advocates of Cornell realism have been struggling to make sense of genuine moral disagreement and account for the purpose of moral inquiry. To circumvent these challenges, David Brink

(1989; 2001; 2007) has proposed an alternative view of causal regulation according to which moral terms refer to whatever complex descriptive properties are needed for interpersonal justification of people's characters, actions, or institutions.

Here we argue that Brink's alternative faces other challenges. We argue that he is committed to anti-realism about truth, which appears to be at odds with moral realism. We furthermore argue that moral explanation requires the use of "thick" moral terms, such as "human" and "person," which are distinct from their thin homophonological counterparts. On the thick reading of "human" and "person," the descriptive properties best suited in interpersonal justification of people's characters, actions or institutions turn out to depend on cultural moral convictions about which living beings have full status as ethical persons.

How then to be a naturalistic moral realist? We argue that the deeper reason the Cornell realists fail is that the sort of naturalism they are attracted to—that is to say, methodological naturalism—is irreconcilable with moral realism. But this failure doesn't require us to abandon realism or naturalism. Ontological ethical naturalism is in fact perfectly reconcilable with moral realism. One view that preserves both ontological ethical naturalism and moral realism is the moral sentimentalism previously defended by one of the authors (Michael Slote). On this view, actions are wrong just when they show the agent as lacking in empathy. What it takes for an agent to be lacking or not lacking in empathy is an a priori matter. This means that moral claims like "adding alluring flavors to e-cigarettes is wrong," "Eichmann was evil" and "neo-Nazism is nefarious" have objective truth-values, thus vindicating moral realism. Once we know the relevant descriptive details surrounding an action, moral sentimentalism holds that determining whether an action is permissible or not is an a priori matter. So, the proposed view entails a rejection of

methodological naturalism and the problems that follow in its wake. However, because the property of showing an agent as lacking in empathy is a natural kind property, moral sentimentalism is a variant of ontological naturalism. So, we conclude, once we give up on the Cornell realist's unrealistic ambition to reconcile two views that are in diametrical opposition, we can indeed have it all.

## 2. CIRCUMVENTING MOORE'S OPEN QUESTION ARGUMENT

In *Principia Ethica*, G. E. Moore presented his now well-known open question argument against moral naturalism, particularly hedonism (Jeremy Bentham 1789), common-good ethics (T. H. Green 1883) and Social Darwinism (Herbert Spencer 1891).<sup>6</sup> According to Moore, if moral properties are natural properties, then moral terms are semantically equivalent to natural kind terms of the sort identified by the natural (and empirical) social sciences. The naturalists of the day were arguing that "what is morally good" is equivalent to "what maximizes pleasure," (Bentham) "psychological fitness," (Spencer) or "what we desire to desire." (Green) But now consider the claim that "*x* is pleasure" is semantically equivalent to "*x* is good." If this claim is true, Moore reasoned, then "pleasure is good" is semantically equivalent to "pleasure is pleasure." Yet while it is not possible to doubt an uninformative tautology like "pleasure is pleasure," it is most certainly possible to doubt a claim like "pleasure is good." Unlike "pleasure is pleasure," "pleasure is good" is not self-evident but remains an open question.

Moore concluded that moral truths are neither identical to (or replaceable by) purely descriptive truths, nor logically derivable from them. Rather, moral properties (or concepts—Moore didn't distinguish between properties and concepts) are non-natural and *sui generis*, which Moore took to mean that they were simple and unanalyzable. In other

words, they are the kinds of properties we today might call “fundamental” or “basic.” Moore’s belief that moral properties are non-natural and *sui generis* is what drove him to conclude that ethics is autonomous (i.e., independent of the natural and empirical social sciences). But while maintaining that ethics is autonomous, Moore insisted that moral facts are nonetheless cognitively accessible to us. Although he at times seemed to hold that intuition is fallible if not unreliable, his considered view was that we come to know moral truths through intuition.

Moore wasn’t merely opposed to ethical naturalism but also to ethical supernaturalism, the view that there is no interesting (i.e., causal, metaphysical or logical) relationship between moral properties and descriptive properties. He argued that moral properties morally supervene on complex descriptive properties, which is to say, two situations cannot differ in their moral properties without differing in their descriptive properties (See Väyrynen 2018). Moral properties are constituted by complex descriptive properties and therefore ontologically depend on complex descriptive properties, but the same moral properties can be realized by different complex descriptive properties.

Moore’s claim that moral truths morally supervene on purely descriptive truths yet are neither identical to (or replaceable by), nor logically derivable from, purely descriptive truths may seem superficially akin to ethical naturalism of the kind defended by Cornell realists. The crucial difference between ethical naturalism à la Cornell realism and non-naturalism à la Moore is that the naturalists explain moral supervenience by appealing to some naturalistically acceptable relation (identity, functional realization, metaphysical grounding, etc.), whereas non-naturalists take moral supervenience to be a brute fact.<sup>7</sup> Non-naturalists are silent about why the dependence relation obtains (Blackburn 1984, 186; Schiffer 1987, pp. 153–154). When

supervenience is explainable, as naturalists maintain, it is also known as “superdupervenience” (a term coined by William Lycan 1986).

Let’s turn now to the question of whether Moore’s open question argument succeeds in refuting ethical naturalism. Today, most thinkers think it misfires. Moore’s argument rests on the claim that moral properties cannot be descriptive properties because one can rationally believe that the descriptive properties are instantiated without believing that the moral properties are instantiated as well. Yet we can only doubt a claim like “pleasure is good” if we are in a position to rationally believe that something is pleasure and yet not rationally believe that it is good.

Moore’s notion of dubiousness is reminiscent of the Fregean notion of cognitive informativeness (or cognitive significance). “Hesperus is Phosphorus” can be cognitively informative, whereas “Hesperus is Hesperus” cannot. If terms contribute descriptive senses to truth-conditions, as Frege thought, this is predictable. “Hesperus” and “Phosphorus” are presumably associated with different descriptive senses; as a result, we can grasp the sense of “Hesperus” without automatically grasping the sense of “Phosphorus.” Similarly, “*x* is pleasure” and “*x* is good” are in all likelihood associated with different descriptive senses, which should make it possible for us to believe that something is pleasurable yet disbelieve that it is good.

As Saul Kripke (1970/1980) convincingly argued, however, Frege’s descriptive senses, or descriptions more generally, should not be mistaken for the *semantic content* (or semantic values) of referential terms. This view is also known as the “neo-Russellian” or “Millian” semantics of directly referential terms, named after its founding fathers Bertrand Russell and J. S. Mill. Various aspects of the semantics have also been advanced by Ruth Barcan Marcus (1961; 1963), Keith

Donnellan (1966), Alvin Plantinga (1969), Peter Geach (1969) and Hilary Putnam (1975).<sup>8</sup>

One of Kripke's arguments against a simple descriptivist account of reference is the so-called "modal argument," which can be summarized as follows. If "Aristotle" were semantically equivalent to the description "the teacher of Alexander the Great," then "Aristotle is the teacher of Alexander the Great" and "the teacher of Alexander the Great is the teacher of Alexander the Great" would have the same modal profile. Yet the latter is true in all possible worlds in which Alexander the Great exists and has a unique teacher, whereas the former is false in those possible worlds where Alexander the Great had a different teacher or where Aristotle entered a different profession.

Advocates of Cornell realism hold that Moore's argument fails for much the same reason that the simple description theory fails. Moore assumes that the cognitive informativeness of, or ability to doubt, an identity claim involving a term A and a term B automatically shows that A and B are not semantically equivalent. But if the meaning of A and B just is their referent, then A and B can be semantically equivalent without this being self-evident. This is, roughly, how Cornell realists like Boyd, Brink and Sturgeon respond to Moore's challenge. Like Kripke and his followers, they invoke a causal theory of reference to explain how the referent of a moral term can come to constitute its meaning. This is also what is supposed to explain why moral properties supervene on descriptive properties. Let's have a quick glance at the details of their view.

### 3. BOYD'S CAUSAL SEMANTIC ACCOUNT

Cornell realism comprises three main theses (Sturgeon 1985; Railton 1986; Boyd 1988; Brink 1989). (i) Moral terms denote moral properties that (ii) are identical to (and hence supervene) on descriptive functional

properties, which (iii) causally regulate our use of the moral terms.

The proposal here is analogous to a fairly standard account of natural kind properties like *being gravity*, *having kinetic energy*, *being water*, *being a DNA molecule*, *having a desire* or *having skin pain* as identical to (and hence supervenient on) descriptive properties that jointly play a particular causal or explanatory role in a given science. For example, the kinetic energy of an object moving at a certain velocity is identical to the energy that will cause the object to accelerate from rest to its current velocity, as described by Newton's laws. Similarly, to have skin pain is identical to a certain stimulation of nociceptors in the skin that causes nerve signals to be transmitted to and activate specific sensory areas in the brain via the spinal cord and brainstem, as described by neuroscience.

In a similar vein, advocates of Cornell realism argue that moral properties like *being right* or *being morally depraved* are identical to (and hence supervene on) some descriptive functional property N that plays a particular causal or explanatory role in our best ethical theory. (e.g., N might be the functional property of maximizing human pleasure, that of maximizing human flourishing or that of always treating others as an end in themselves and not as a mere means to an end.) These descriptive functional properties (e.g., the functional property of maximizing human pleasure) furthermore supervene on their realizers (the properties that fulfil the "moral"-role).

To explain the dependence of moral properties on purely descriptive properties, Cornell realists take inspiration from the causal theory of reference, the metasemantics made famous by Kripke (1970/1980) and Putnam (1975). On this view, the use of a referential term, such as a proper name or a kind term, refers to whatever the term is appropriately causally connected to. The referent is initially fixed by a speaker using whatever means will allow

her to talk about a unique thing or kind of thing. For example, you can fix the reference of “Foxtrot” by pointing to your new puppy, saying “He is to be called ‘Foxtrot’.” Or, assuming there is a ninth planet, an astronomer can fix the reference of the name “Planet Nine” by saying: “The ninth planet in our solar system is to be called ‘Planet Nine’.” A speaker can even single out a unique person using a description that doesn’t accurately describe that person. To use a Donnellan-esque example (Donnellan 1966): at a party, you spot your friend Walter at the other side of the room drinking out of a Scotch glass. You turn to your coworker and say: “Look, the man drinking Scotch is my old friend Walter.” If Walter is drinking Ginger Ale from a Scotch glass, the description is inaccurate, but as long as your coworker can figure out who you are talking about, inaccuracy doesn’t matter.

The reference fixing of natural kind terms follows a similar pattern. For example, a long time ago someone might have said “Let the word ‘water’ denote the potable liquid that fills wells, rivers, and lakes, quenches our thirst and keeps us floating when we swim.”

Once the reference of a referring term is fixed, the referring term is then passed on from speaker to speaker through communication. This is what is also sometimes called “reference borrowing” or “reference inheritance.” Many years later, when the name has been passed on countless times, speakers can still use it to refer to the thing or kind of thing that was initially dubbed with that name, provided that their uses of the referential term are links in an unbroken causal-historical chain stretching back to the initial dubbing.

Advocates of Cornell realism have extended the causal theory of reference to moral terms in different ways, but Richard Boyd’s (1988) approach is commonly taken to be representative. According to Boyd, moral terms, like “right” and “wrong,” are natural kind terms that refer to the descriptive functional

property N (whatever it is) that connects the user of the moral term to N in an appropriate causal-historical chain terminating in its very first use. Following Boyd, we can then say that N causally regulates speakers’ use of the moral term.

Since natural kind terms designate the same properties in all possible worlds in which those properties exist, they are rigid designators. As moral terms are natural kind terms, they too are rigid designators. So, Boyd is committed to the following thesis (cf. Timmons 1999).

*Boyd’s Causal Regulation Thesis*

A moral term T rigidly designates the purely descriptive property, N, that causally regulates our uses of T.

The causal regulation thesis explains why the semantic values (or semantic contents) of moral terms supervene on descriptive properties. The supervenience relation obtains, because the descriptive functional properties just are the semantic values of the moral terms. So, Boyd’s ultimate view can be expressed as follows:

*Boyd’s Naturalism*

Moral property M expressed by moral term T supervenes on descriptive property N by virtue of N’s causal regulation of our uses of T.

Boyd’s strategy for blocking Moore’s open question argument is to replace his presupposed descriptivist semantics with the causal theory of reference. The causal theory makes it possible for the terms of an identity claim to be semantically equivalent, even when we doubt its truth. For example, “Hesperus is Phosphorus” is cognitively informative, which makes it possible to doubt its truth. Yet “Hesperus” and “Phosphorus” refer to the same thing, namely the planet Venus. So, “Hesperus,” “Phosphorus,” and “Venus” all have the same meaning and reference, which is to say that “Hesperus is Phosphorus” is semantic equivalent to “Hesperus is

Hesperus” and to “Venus is Venus.” Similarly, Boyd argues, while we may well doubt that “*x* is pleasure” and “*x* is good” have the same meaning, the causal theory leaves it open whether “*x* is pleasure” and “*x* is good” are semantically equivalent. Problem solved.

#### 4. MORAL TWIN-EARTH AND THE OBSOLESCENCE OF MORAL INQUIRY: OBJECTIONS TO BOYD’S NATURALISM

The perhaps best-known objections that have been mounted against Boyd’s appeal to the causal theory to circumvent Moorean challenges are what we might call the “objection from the obsolescence of moral inquiry” and the “Moral Twin Earth objection.”

David O. Brink (2001) credits the objection from the obsolescence of moral inquiry to John G. Bennett (in conversation). He writes:

Stated baldly, the worry is that substantive moral reasoning and theorizing become obsolete if we let the ethical naturalist respond to the [Open Question Argument] by appeal to the causal theory of reference. The causal theory of reference appears to make the reference of speakers’ use of moral predicates a matter of empirical—specifically, historical—fact insofar as subsequent use of these predicates inherits reference, via speakers’ intentions to use the predicates with the same reference as earlier speakers, from the original use of those predicates to name interesting features of the environment of those original speakers. If so, it would seem that the causal theory implies that disputes about the meaning or reference of moral terms ought to be resolved not, as one would think, by moral reasoning, but by means of a historical inquiry about which features of actions, people, and institutions moral appraisers intended to pick out when those moral terms were introduced. If so, appeal to the causal theory of reference is not a good response to the [Open Question Argument]. (2001, p. 164)

Brink is right that one way to fix a referent, according to the classic Kripke-picture, is through some, conscious and deliberate, referential intention (a “historical fact”). In

such cases, the fields of history and linguistics would be of some use. This might make substantive moral reasoning and theorizing obsolete. But this doesn’t seem to present a problem for Boyd, as reference-fixing seldom is through some, conscious and deliberate referential intention. Boyd’s theory makes substantial empirical assumptions about the nature of moral properties. These assumptions point towards the idea that the initial dubbing of moral properties happens unconsciously. Roughly, Boyd argues that what regulates the use of moral terms is a cluster of natural/biological needs (1988, 4.3.). In this sense, there was never a conscious and deliberate initial baptism. Rather, in virtue of our biological make-up, we are inclined (unconsciously, most of the time) to track those properties whose instantiation satisfies our natural needs. But figuring out the true nature of our needs is a substantial first-order ethical question.

The Moral Twin Earth objection, which is due to Mark Timmons and Terence Horgan (1991; 1992a; 1992b; Timmons 1999), seems to present greater difficulties for Boyd’s theory.<sup>9</sup> In his original Twin Earth thought experiment, Hilary Putnam (1975) invites us to imagine that in 1750, there was a distant planet, Twin Earth, which was qualitatively indistinguishable from Earth except that the watery stuff that flows in rivers, lakes and oceans is composed of the chemical XYZ instead of H<sub>2</sub>O. XYZ is superficially indistinguishable from water: it looks and tastes like water, is drawn from the well and can be used to open up your Scotch whisky. Back in 1750 scientists on Earth had not yet discovered the molecular structure of water. Similarly, scientists on Twin Earth had not discovered the composition of Twin-water. But, if we reflect on how we use the word “water,” it should be clear that it would have referred to H<sub>2</sub>O and not XYZ, even when uttered by an earthling in 1750. Conversely, if an inhabitant on Twin Earth had used the word “water” in 1750, she would have referred to XYZ and not H<sub>2</sub>O.

In Timmons and Horgan's Moral Twin Earth thought experiment, Earth and Twin Earth are exactly alike, except that the people on the two planets use different standards for moral assessment (Horgan & Timmons 1991; 1992a; 1992b; Timmons 1999). Earthlings abide strictly by Bentham's hedonism, whereas inhabitants of Twin Earth are unrelenting Kantians. Suppose now that Oscar, an Earthling, meets Twin-Oscar, a Twin-Earthling. Twin-Oscar runs the following Dictator scenario by Oscar. "Suppose a dictator is going to shoot ten people lined up in front of you, unless you shoot one of them. You make a random choice and shoot one, thereby saving nine lives. Was your choice right or wrong?" Being a committed hedonist, Oscar replies that the choice was right. Twin-Oscar, committed as she is to Kant's Principle of Respect, denies that it was the right choice.

According to Timmons and Horgan, it seems that Oscar and Twin-Oscar are engaged in genuine moral disagreement. But here is the problem. The causal theory of reference commits us to the view that the Earthian term "right" and the Twin-Earthian term "right" have different meanings. If this is so, then the dispute between Oscar and Twin-Oscar is not a genuine moral dispute after all. Rather, it's theoretically related to a verbal dispute between two people who use the ambiguous word "bank" to mean different things, as in the following discourse fragment: Al says: "Cy told us to meet him at the bank [financial institution]." Bo replies: "No, he definitely didn't tell us to meet him at the bank [river bank]." Boyd's semantics of moral terms, Timmons and Horgan conclude, is mistaken, because it's unable to account for the intuition that Oscar and Twin-Oscar are engaged in genuine moral disagreement.

##### 5. BRINK'S CAUSAL SEMANTICS

The philosophical literature is spawned with responses to the Moral Twin Earth argument ranging from identifying significant

disanalogies between Putnam's thought experiment and Moral Twin Earth (Laurence et al. 1999) to denying its metaphysical possibility (Kraemer 1990, 468) or the reliability of our intuitions about the case (Sonderholm 2013).

David Brink (2001), one of the original Cornell realists, has offered a rather different response to Moral Twin Earth. To bring to the fore what's bothersome about Timmons and Horgan's objection, Brink asks us to imagine an intra-planetary version of Moral Twin Earth à la the following. Suppose here on Earth, Oscar studied in a Hedonist country, whereas Twin-Oscar studied in a Kantian country. They meet and disagree about the Dictator scenario. If their disagreement is not genuine, because Oscar's use of the word "right" refers to the property of maximizing human pleasure, whereas Twin-Oscar's use of the word "right" refers to the property of always treating others as an end in themselves and not as a mere means to an end, then the meaning and referent of moral terms depend on local moral traditions. Cornell realism, it would seem, turns out to be a version of moral relativism.

Brink's solution is two-fold. (i) He offers an account of the meaning of moral terms that allows people engaged in genuine moral inquiry to disagree about a shared content, even if they have different moral convictions. (ii) He rejects Boyd's extensional approach to causal regulation in favor of an epistemic account of causal regulation.

Let's begin with (i). Brink draws a distinction between concepts and conceptions of morality. Hedonism and Kantian deontology are different conceptions of morality. But they nonetheless still share a common (abstract) concept of morality. This shared common concept of morality allows people with different moral convictions to engage in moral debate about what the best conception of the concept of morality is. Anyone who genuinely engages in moral debate or inquiry,



Brink argues, does so with the intention of picking out the grounding moral properties (e.g., the property of maximizing human pleasure) that will meet standards others can and should accept. As he puts it:

On this view, we should understand perhaps all moral appraisers, and certainly those who introduced moral categories and terms, as using those categories and terms with the intention of picking out properties of people, actions, and institutions—whatever those properties are—that play an important role in the interpersonal justification of people’s characters, their actions, and their institutions. (2001, p. 174)

As anyone who is genuinely engaged in moral debate or inquiry must adopt this interpersonal justificatory concept of morality, we can also simply think of the concept as “the moral point of view.” If people don’t adopt this point of view when they assess people, actions or institutions, then they are not engaged in genuine moral assessment. Thus, when Oscar, the hedonist, and Twin-Oscar, the Kantian, disagree about whether to shoot in the Dictator scenario, their disagreement can be loosely paraphrased as follows (let *S* be the action of shooting one prisoner):

*Oscar:* *S* has the property needed for the interpersonal justification of people’s characters and actions.

*Twin-Oscar:* *S* does not have the property needed for the interpersonal justification of people’s characters and actions.

This raises the question of how to reconcile Brink’s proposal and the neo-Russellian/Millian account of the semantic content of natural kind terms. On the latter view, the semantic content of a name or natural kind term just is its referent. But Oscar uses “right” to refer to the property of maximizing human pleasure, whereas Twin-Oscar uses “right” to refer to the property of always treating others as an end in themselves and not as a mere means to an end. So, if the referent doubles as the semantic content, then Oscar and Twin-Oscar

do not disagree about a shared semantic content.

It may be thought that one can reconcile Brink’s proposal with a neo-Russellian/Millian semantics if one takes the shared content in moral disputes to be linguistic meaning or “character” in David Kaplan’s (1989) sense, viz. a function from the context of utterance to semantic content.

However, this move would be unwise. When the linguistic meaning of a disambiguated term doesn’t suffice to fix its semantic content, the parameters of the context of utterance together with the term’s linguistic meaning fix its semantic content, which is to say that the term is context-dependent. Indexicals and demonstratives are prime examples. The linguistic meanings of “I,” “now,” and “here” don’t fix their semantic contents. What fixes their contents are their linguistic meanings plus a context of utterance. For example, If Brink says “I am here now” in San Diego at 3 p.m. on October 2, 2019, Standard Pacific Time, the semantic content is: *Brink is in San Diego at 3 p.m. on October 2, 2019, Standard Pacific Time.* But assuming the linguistic meaning of “right” is *the property needed for interpersonal justification*, then the question arises which parameter of the context of utterance fixes its semantic content. None of the standard Kaplan parameters (world, time, location, speaker, hearer) will do. We would need to add a new contextual parameter, such as “the speaker’s moral standards” or “the moral standards of the speaker’s community.” On this proposal, when Oscar, the hedonist, uses the word “right,” its semantic content is fixed by *his* moral standards and the linguistic meaning of “right.” So, the semantic content of “right” in his mouth is the property of maximizing human pleasure. But there is a problem. This is not moral realism but moral contextualism, or what Gilbert Harman dubbed “moral relativism.” (Harman 1975; Dreier 1990).<sup>10</sup> A more recent variant of moral relativism takes the semantic

content of moral terms to be context-invariant properties, for example, functions from centered worlds to extensions (Brogaard 2007). But this evidently isn't moral realism either, as the truth-values of moral judgments vary from speaker to speaker. So, this sort of move doesn't get us anywhere. And it isn't the path taken by Brink.

Brink rejects the standard causal theory of reference, employed by Boyd. This move is further motivated by Brink's desire to avoid the absurd implication that historians, linguists, and anthropologists should be the authorities on what we mean by our moral terms. In its place, Brink offers an epistemic theory of causal regulation that is centered around the notion of dialectical equilibrium (an idea derived from Rawls' reflective equilibrium; Rawls 1971, pp. 19–21, 46–51, 578–581). Here is how it works. Particular moral claims, such as the claim that it would be wrong for Al to break his promise to Bo, are made right or wrong by general rules, such as the rule that one ought to keep one's promises. Rules, in turn, are made right or wrong by ethical principles such as Hedonism or Kantian deontology, although there may be a plurality of basic principles rather than just a single master principle.

Unfortunately, this ontological dependence does not reflect how we attain knowledge of what is right and wrong, as ontological dependence doesn't imply epistemic dependence. While we might sometimes resolve disputes about particular moral claims by finding a general ethical principle that we agree on, it could also be that we agree on particular claims and use those to resolve disputes about ethical principles. For example, we might intuit that it would be wrong for Al to break his promise to Bo. We can then test our ethical principles against the particular claims we agree on. This is what Brink calls the method of "dialectical equilibrium" (Brink 2001; 2007).

Ideally, we modify our moral convictions in response to conflicts until our ethical views

are in dialectical equilibrium. Dialectical equilibrium, however, is an ideal that we can hope to approximate but cannot meet. In order to approach this ideal, we will likely need to revise some of our current moral beliefs, but—Brink argues—there is no way of knowing in advance just how revisionary our moral beliefs would be, were they in dialectical equilibrium.

## 6. BRINK'S "ANTI-REALIST" REALISM

Let's now turn to our first concern about Brink's proposed fix. If we focus narrowly (for now) on the moral terms "right" and "wrong" as applied to human action, Brink's causal regulation thesis can be articulated as follows:

### *Brink's Causal Regulation Thesis*

Descriptive property N causally regulates a speaker S's use of moral term T if and only if for any action A, S's application of T to A would be dependent on her belief that A has N, were S's beliefs in dialectical equilibrium.

Since reference is logically tied to truth, Brink's thesis entails a theory about ethical truth. Where descriptive property N causally regulates S's use of the moral term "right," and S believes that A is N (for some N, S and A), Boyd's thesis implies: A is right if and only if, if S's beliefs were in dialectical equilibrium, then S would believe that A is right (we leave the proof to the reader). Let's simplify. We will use the following abbreviations:

$Bp$ : S believes that  $p$ .

$\forall$ : Universal quantifier.

$p \boxtimes \rightarrow q$ : If  $p$  were the case, then  $q$  would be the case (the David Lewis counterfactual).

$p \Leftrightarrow q$ : Necessarily,  $p$  if and only if  $q$  (the necessary material, or strict, biconditional).

$Q_N$ : For some agent S, S's beliefs are in dialectical equilibrium in N-circumstances.

N-circumstances: Circumstances in which N regulates S's use of "right," and S believes that A is N.

Brink's thesis about ethical reference thus implies the following thesis about moral truth:

*Moral Truth*

$$\forall A (A \text{ is right} \Leftrightarrow (Q_N \boxtimes \rightarrow B(A \text{ is right})))$$

Interestingly, this result (*Moral Truth*) is just the moral version of anti-realism about truth. Here is Alvin Plantinga's formulation of anti-realism about truth from his presidential address to the APA, "How to be an Anti-realist" (1982, 64–66): *p* is true just in case if ideal circumstances were to obtain, then a suitably placed person would believe that *p*. Or:

*Anti-Realism about Truth*

$$\forall p (p \Leftrightarrow (Q \boxtimes \rightarrow B(p)))$$

In plain English: truth just is that which would be believed in ideal circumstances. Where descriptive property N causally regulates S's use of "right," and S believes A is N (for some S, N and A), substituting "A is right" for "*p*" and " $Q_N$ " for "*Q*" yields the above principle *Moral Truth*—a principle to which Brink is committed. In fact, Brink's general thesis entails anti-realism about moral truths in general, including truths about the goodness or badness of agents and institutions and truths about ethical principles and rules.

So, moral realism à la Brink is an instance of anti-realism about truth. Or so it would seem. As an anonymous reviewer has pointed out to us, Brink might be willing to accept the biconditional that we attribute to him, yet not take it to define "moral truth." According to the reviewer, there could be, say, a joint third-factor explanation for why the two sides of the strict conditional obtain.

Whether this is a realistic response remains to be seen. But we are skeptical. It is true, of course, that the two sides of a strict conditional truth needn't be conceptually (or a priori) equivalent. Although "(*x* is an equilateral

triangle)  $\Leftrightarrow$  (*x* is an equiangular triangle)" is true, it doesn't follow that "equilateral" and "equiangular" are interdefinable. Still, if Brink's view entails anti-realism about moral truth, then it's hardly a form of realism.

But there is more: even when restricted to moral truths, anti-realism about truth has odd consequences. How odd the consequences are depends on how inclusive the domain of moral propositions is. We already know that the domain of moral propositions includes propositions about specific actions like *Al should keep his promise to Bo*, propositions about rules such as *people should keep their promises* and propositions about ethical principles such as *hedonism is true*. What about a proposition like *For some S, S's beliefs are in dialectical equilibrium in N-circumstances*? The answer is that Brink is required to say that this proposition doesn't fall into the domain of the ethical. For suppose otherwise (recall that  $Q_N$  stands for: *For some agent S, S's beliefs are in dialectical equilibrium in N-circumstances*). The following is then an instance of anti-realism about moral truth): ( $Q_N \Leftrightarrow (Q_N \boxtimes \rightarrow B(Q_N))$ ). But this entails  $Q_N$  (see the Appendix for the proof).<sup>11</sup> Or in plain English: If propositions like *For some agent S, S's beliefs are in dialectical equilibrium in N-circumstances* are included in the domain of the moral, then anti-realism about moral truths entails that it is *necessary* that for some (properly placed) agent S, S's beliefs are in dialectical equilibrium in circumstances in which descriptive property N regulates S's use of "right," and S believes that action A is N. So, anti-realism about moral truths implies that *necessarily* there is a properly placed epistemic agent. Since human placement in epistemic circumstances (not to mention human existence itself) is a contingent matter, the properly placed epistemic agent cannot be human but must be superhuman or a divine agent, such as God. So, without domain restrictions, anti-realism about moral truth entails supernaturalism. Not a satisfactory

result. This is basically what Alvin Plantinga concluded in his 1982 presidential address to the APA, “How to be an Anti-realist.” He argued that an anti-realist about truth should be a theist (see Brogaard 2016).

Brink explicitly rejects supernaturalism, stating that “so far from our knowledge of God’s will supplying evidence about the nature of morality, it is our beliefs about the nature of morality that are supplying evidence about God’s will” (2007, p. 159).

To avoid substantial theistic consequences, Brink can (and must) deny that propositions like *S’s beliefs are in dialectical equilibrium in N-circumstances* fall under the domain of moral propositions. But there is another problem in the vicinity. Brink states that “dialectical equilibrium is an ideal that none of us now meets and [that] we can at most approximate” (2007, p. 156).

This is a peculiar claim. Given Brink’s rejection of supernaturalism, the relevant epistemic agents must be human. But if dialectical equilibrium is humanly unattainable, as Brink seems to think, then it is unclear what explanatory advantage his variant of Cornell realism has over supernaturalism. With the supernaturalist, Brink would be forced to acknowledge that moral truth cannot be understood in terms of *our* epistemic capabilities.

Perhaps there is a way to ward off this objection. But, as we saw above, Brink is still settled with moral anti-realism, which makes his view ill-suited to serve as a variant of realism. In the next section, we will present an informal argument to show that Brink’s view entails anti-realism in a form that is considerably harder to swallow than Plantinga’s anti-realism.

## 7. CORNELL REALISM AND RELATIVISM: THICK AS THIEVES

In the previous section, we argued that Brink’s naturalism is necessarily equivalent to anti-realism about moral truth. We will now

present an informal argument to show that Brink’s view entails anti-realism in a form that is considerably harder to swallow than Plantinga’s anti-realism. Our argument rests on the common distinction between thin and thick moral properties, or concepts.<sup>12</sup>

*Lewdness, selfishness, dishonesty, generosity, modesty, and altruism* are prime examples of thick ethical concepts. Thick ethical concepts contrast with thin ethical concepts, such as *good, right, impermissible, obligatory, and virtuous*, which connote only an evaluative component. The distinction dates back to Bernard Williams (1985), who argued that thick concepts hold together a descriptive component and an evaluation, where the evaluation is either *pro* or *con*. Aristotelian virtues such as *honesty, courage, and temperance* are thick *pro*-concepts, whereas vices such as *cowardice, insensibility, injustice, and vanity* are thick *con*-concepts. *Honesty*, for example, holds together the descriptive component of telling the truth when and only when the situation calls for this behavior and the positive evaluation of possessing this character trait.

After these preliminary remarks, let’s turn to our argument. Brink argues that the meaning and reference of “thin” moral terms, like “good,” “bad,” “right,” “wrong,” “virtue” and “vice,” etc., ought to be determined through substantive moral dialectical reasoning and theorizing rather than, say, through historical inquiry about how thin moral terms were used historically. But substantive moral reasoning and theorizing involve moral explanations, and at least some moral explanations involve the use of thick moral terms. Suppose your teen daughter wants to know why her brother is grounded. Saying “He did something wrong” is not as explanatorily satisfactory as saying “He was dishonest when I asked him why I could smell smoke on his breath.” Here, *dishonesty* is a thick moral concept. But some thick moral concepts will inevitably depend on the beliefs of agents or communities.

To see this, imagine that your philosophy colleagues Kate and Carl disagree about whether our so-called fundamental human rights to life and liberty are inalienable. Kate argues that they are. Qua inalienable, they cannot be lost under any circumstances. Carl rejects the very idea of *inalienable* human rights. According to him, most of us have the rights to life and liberty, but those rights are not inalienable. Much like privileges they can be lost. Individuals who commit extremely cruel acts as a mere means to their own selfish ends automatically lose those rights, that is to say, their cruelty disqualifies them as counting as humans (or human persons).

As employed in the disagreement between Kate and Carl, the philosophical term “human” (or “person”) is a thick moral term. Why think this? Well, Kate and Carl agree that the scientific term “human being,” or “*Homo sapiens*,” applies to all of us, even the most cruel among us. But they disagree about whether the same goes for the morally thick term “human.”

The morally thick term “human” thus turns out to have a different meaning and reference in the mouths of Kate and Carl. Unlike Carl, Kate uses the terms “human” and “*Homo sapiens*” to refer to the very same individuals. As meaning determines reference, and they agree on the meaning and reference of “*Homo sapiens*,” the term “human” does not have the same meaning and reference in the mouths of Kate and Carl.

But moral dialectical reasoning and theorizing can be meaningful or substantive only if the thick moral terms employed by the interlocutors have the same meaning and reference. As “human” doesn’t have the same meaning and reference in the mouths of Kate and Carl, their debate isn’t substantive and therefore doesn’t get them any closer to dialectical equilibrium.

The envisaged example of Kate and Carl can be extended to whole communities. It’s not hard to imagine that a term like “human”

(or the equivalent in other languages) could acquire different meanings and denotations fixed by different original intentions to adopt the moral point of view. Imagine a world, call it “Kate-Land,” in which the dominant moral opinions concur with Kate’s moral view, whereas the dominant moral views in another world called “Carl-County” concur with Carl’s views. The fact that this sort of scenario is morally possible runs counter to Brink’s commitment to the idea that moral terms have their reference fixed by a single original intention to adopt the moral point of view, that is, Brink’s claim that moral terms refer to an interpersonally justifiable construct appears to be indefensible.

But what we have said here about thick moral terms carries over to thin moral terms. “The death penalty is a morally wrong institution” is true in Kate-Land but false in Carl-County, where the dominant view is that extremely cruel people fall short of being humans or persons with inherent dignity, despite being members of our species.

The upshot is this: the meaning and reference of both thin and thick moral terms are different in Kate-Land and Carl-County. So, the meanings and referents of thick and thin moral terms alike depend on historical events of a given culture and the resulting local moral conventions, which is to say, Brink’s naturalism isn’t a form of realism.

Before proceeding, let’s consider a couple of objections. Perhaps Brink could argue that the people in Carl-County haven’t reached or gotten close enough to dialectical equilibrium. But on what grounds? Brink states that, on his moral realist view:

[W]e should understand perhaps all moral appraisers, and certainly those who introduced moral categories and terms, as using those categories and terms with the intention of picking out properties of people, actions, and institutions—whatever those properties are—that play an important role in the interpersonal justification of people’s characters, their actions,

and their institutions. Subsequent appraisers inherit this intention, if only because they use the same words as their predecessors and have the intention of continuing an inquiry into the same subject. (2001, p. 174)

In Carl-County, morally thick terms like “human person” and morally thin terms like “right” and “good” refer to whatever complex descriptive properties are needed for *interpersonal* justification of people’s characters, their actions, or their institutions. As the morally thick term “human” (or “person”) does not apply to cruel individuals in Carl-County, “human” refers to a descriptive property that is coextensive with that of being an individual whose behavior is not cruel by any measure. This property satisfies Brink’s requirement that a property is genuinely moral only if it is needed for the *interpersonal* justification of people’s characters, their actions, and their institutions. The same, of course, can be said about Kate-Land.

Perhaps Brink may insist that the conditions under which it is reasonable to interpret people’s judgments as moral judgments do not obtain in, say, Carl-County. Unfortunately, he cannot avail himself of this strategy. This is because Brink is committed to the view that said conditions obtain as long as the judgments of the people are based on standards that are generally endorsed as interpersonally justifiable. He writes:

Even when appraisers from distinct communities use language in ways that are counterfactually regulated by different properties of their environments, we should interpret their language as moral language and the judgments that employ that language as moral judgments only if those judgments are based on standards that the appraisers endorse, if only implicitly, as interpersonally justifiable (2001, p. 174).

The vast majority of residents of Carl-County endorse standards that involve morally condoning the death penalty as an *interpersonally* justifiable institution, whereas

the vast majority of residents in Kate-Land endorse standard that condemn the death penalty as an interpersonally justifiable institution. The upshot: not only is Brink committed to anti-realism about truth, but he is also seemingly committed to Harman-style moral relativism that rejects the idea of objective moral truths. In the end, then, Brink may be even worse off than Boyd.

## 8. HOW TO BE A NATURALIST

The foregoing casts doubt on the Cornell School’s naturalistic moral realism and on David Brink’s attempts to improve on its assumptions and methods. But there is an objection to naturalism that doesn’t surface in the previous discussion and doesn’t seem to have been considered by the philosophers who have so far been criticized here. The objection has been put most explicitly and forcefully in recent years by Sharon Street. Street is perhaps best known for her arguments, based on evolutionary considerations, (mainly) against recent ethical rationalists like T. M. Scanlon (Street 2006; Scanlon 1998). Those arguments seek to show that evolution cannot plausibly be thought to have put us humans in touch with the non-natural moral truths that these neo-Kantians widely subscribe to. Moreover, these moral realists rely on our ability to grasp and possess reasons that, ontologically, exist outside the natural realm, and this seems a strike against Kantian moral realism and in favor of a naturalism that makes no such assumptions.

However, Street thinks the naturalists have another and different problem. The Kantians can claim that our grasp of reasons explains our moral or other motivation. If we see we have a reason to do something, that can plausibly be thought to already involve at least some motivation to do it. So, the neo-Kantian picture to be found in Scanlon and other contemporary rationalists explains why moral (or prudential) considerations have rational and motivational force. It may be based on

initially suspect ideas about the non-natural realm, but at least it provides for and explains the *normativity* of such considerations, and that has long been considered (though not by everyone: e.g., Brink 1989) a desideratum for any philosophical account of morality. By contrast, Street (2008) argues, ethical naturalism appears to be unable to account for moral normativity and so fails in its own different way to give us what we want from a philosophical account or theory of morality. According to Street, non-naturalist ethical rationalism is an inadequate form of moral realism because it doesn't fit comfortably with what we know about evolution and the evolution of moral norms and values, but naturalism isn't "worth worrying about" as a form of moral realism because it fails to do justice to the normativity of the moral.

If you look at the most well-known forms of naturalism, Street seems to have a point. Views that see morality as a matter of what is more rather than less evolved or (more frequently, of course, nowadays) as a matter of what promotes the welfare/happiness of humanity or sentient beings generally have a difficult time explaining why such considerations are inherently motivating. Knowing that some action will help humanity generally, why does that automatically make someone want to perform that action? Why shouldn't they instead be totally interested in doing what will help themselves or their families? The point can be made more easily in regard to evolutionary ethics, and it is difficult in general to see how any viable form of naturalism could meet this test or, then, the test of automatically providing a reason or reasons for action. Ethical rationalism tells us we are in touch with reasons, non-natural reasons, motivating us toward, e.g., helping others, but the kind of ethical naturalism we are familiar with cannot offer this kind of foundation for normativity, and all things considered that represents a problem for naturalism as a basis for moral realism.

A few ancillary clarifications are needed at this point. Some moral realists think moral judgments don't have to involve normativity and can be purely descriptive. They therefore wouldn't consider the just-mentioned objection to be an objection to their version of moral realism (e.g., Brink 1989). But wouldn't it be philosophically preferable if we could account for normativity within a realist framework that didn't have to appeal to extra-natural factors? In what follows we are going to show you how we think we may be able to accomplish that. The argument will make central and essential use of the notion of empathy, something the Cornell School never thought of doing. But first some further clarifications.

Ethical egoism of the sort that treats our own pleasure or contentment as our ultimate moral goal is a moral realist form of naturalism, but no one or almost no one nowadays regards egoism as a moral theory rather than, at best, a theory of what it is rational to do. Egoism imposes too high a philosophical and moral price for someone seeking a naturalist form of moral realism to pay. Then, too, there are forms of subjectivism, non-cognitivism, and error theory that make no mention of anything beyond the natural, and the first of these also allows for moral truth. Subjectivism could conceivably be viewed as a form of naturalistic moral realism, but it involves a kind of relativism that (as per the first parts of this paper) is not what philosophers who call themselves moral realists are looking for.

We must look elsewhere, and we propose that we look toward or into the possibilities of moral sentimentalism. This will seem immediately suspect to many of you because the most familiar forms of sentimentalism don't allow for moral realism in the strictest or fullest sense. Sentimentalists holding an error theory of moral attributions or subscribing to some kind of non-cognitivism are obviously not moral realists, but neither, in the fullest sense, are the ideal-observer or

response-dependent theories that some moral sentimentalists subscribe to. Yes, they allow for moral truth and a moral objectivity that holds in relation to human nature, but many of us would consider that to be a problem because we are looking for an account of moral judgments like “it is wrong to torture babies” that regards them as valid independently of specifically human attitudes and dispositions. And unlike the non-naturalist Kantian moral realists, the ideal-observer and response-dependent theorists have entirely given up on that goal. However, that goal is central to Slote’s approach to moral sentimentalism, and the book *Moral Sentimentalism* (MS) was written largely in furtherance of it (Slote 2010). However, one might well wonder whether, when moral sentimentalism (somehow) turns toward defending a fully realist view of morality, its naturalism won’t prevent it from achieving our other (self-given but widely accepted) desideratum, showing how moral considerations and claims can be normative. Let us see.

Moral sentimentalism in the hands of its greatest practitioner, David Hume, relied heavily on empathy, which he called “sympathy” because the term “empathy” didn’t yet exist. (There are times when Hume uses “sympathy” to refer to what we now call “sympathy” and other times when it clearly refers to what we would now call “empathy.”) Since empathy of the kind Bill Clinton was talking about when he said “I feel your pain” involves taking in the emotions, feelings, attitudes of others, it is clear how an emphasis on empathy leaves us fully within the ambit of naturalistic moral sentimentalism. But how is all this compatible with moral realism? Hume in effect didn’t think it was, but there are aspects of empathy Hume never considered, most particularly its cognitive side. Nowadays, many of us think that empathy is a quick and non-inferential way of being acquainted with and aware of how another person is feeling, and Hume never acknowledged (sic)

that empathy could work that way. However, the fact that empathy has a cognitive side to it allows it to bridge the gap between fact and value, knowledge and motivation, in a way that allows a sentimentalism based in empathy to articulate and defend a form of sentimental moral realism and to do so in a way that permits of moral (and other) normativity.

But how exactly does empathy bridge that gap? It is not difficult to see how receptive or emotional empathy represents a kind of acquaintance with another’s feelings or attitudes. But why should such acquaintance be tied to motivation to act (much less reasons to act)? The mostly psychological literature on empathy treats it as an entirely contingent causal/developmental matter that for many or most humans empathy with the suffering or distress of another person makes them want to help that person.<sup>13</sup> However, if empathic knowledge of the other doesn’t automatically involve motivation to help, then empathy lacks normative force and any moral sentimental philosophy based on empathy will seem to be incapable of explaining how moral considerations (based in empathy) can be normative. We seem to be back with our original problems, but are we?

What if the connection between empathy with feelings/attitudes and motivation to act isn’t causally contingent? To be sure, those who have written on these issues—both the psychologists and the philosophers—say or presuppose that the relationship is a contingent one,<sup>14</sup> but in recent years one of us has discovered what we believe to be a sound argument to show that it is conceptual and necessary rather than causal and contingent. The argument is deployed in much of Slote’s recent work going back as far as Slote (2014), but we shall briefly state it again here.

When you as a parent are infected by, say, your daughter’s enthusiasm for stamp collecting, the process can occur without your initially knowing that it is occurring. Empathy



is clearly at work here, but consider that in such a case what you empathically take in is not mere enthusiasm, some unfocused positive state of feeling. No, what you take in is enthusiasm for stamp collecting. You take in a feeling with its intentional object, and fully empathic processes precisely involve both feeling and its target.

So, consider a case where someone is distressed by the sharp pain in their arm. *Ex vi termini* distress means wanting to get rid of or diminish what one is distressed about, so the distressed person wants to get rid of or diminish the pain in their arm. But then if you empathically take in a person's distress *at the pain in their arm*, you too will be distressed by that pain and, again *ex vi termini*, you will be motivated to lessen or do away with that pain, *their pain*. So, with standard empathy as involving the taking in of mental states with intentional objects, empathy necessarily involves motivation to help someone in pain distress or other similar states. To be sure, if someone attacks you with a knife, that motivation to help and the empathy itself may terminate in favor of an overwhelming motivational occupation with self-defense. Also, there is such a thing as "empathic over-arousal" that can lead someone to stop empathically focusing on the other person and instead pay attention to their own unpleasant overload of feeling. But this doesn't affect the main point: that empathy is *necessarily* connected with altruistic desire to relieve the suffering, say, of another person. It would seem, then, that a moral philosophy grounded in empathy has the resources to explain the motivational normativity of moral considerations like the fact that someone else is suffering or in distress.

Ah, you will say, that normativity comes at a price! Even if empathy is more tightly connected with motivation to help than has been realized (even by Hume, who didn't believe in many necessary connections), the sentimentalist who advocates concern with

the suffering of others and sees such concern (as Hume did) as based on empathy seems to have no way to argue that such concern and a morality based on it can avoid non-cognitivism and error theory and possess objective rather than human-relative validity. The emphasis on empathy would seem to box us into a corner where moral realism has to be denied. Moral claims and considerations might be automatically motivating, but how, on such views, would they allow for genuine and in the relevant sense objective reasons to act morally? How, moreover, could such views allow for any sort of moral realism about the truth or falsity of moral utterances or claims? Let us see.

The most important question to consider is whether and how any form of moral sentimentalism can aspire to being a form of moral realism that allows or mandates not only automatic moral motivation but genuine reasons to act morally. The main purpose of *Moral Sentimentalism* was to show that objectively and a priori true moral judgments can be allowed for within a sentimentalist framework, something previous sentimentalists never sought to do. But Slote's point of view, in writing, was that it stands in favor of ethical rationalism that it seeks to show that our ordinary strong moral beliefs (like that it is wrong to torture babies) aren't just expressions or descriptions of human feeling but have an a priori validity. That book adopted a naturalistic conception of moral truth without saying that it was doing so, and it also never took up the issues of normativity and of reasons for action that have been brought into the present discussion. Slote has done some of this in a separate paper subsequent to the book, but we would like to bring those ideas together here, even if only in a sketchy way (that can be amplified by consulting the paper [Slote 2019] just mentioned). So, our first task is to say something in defense of the new idea of a fully realist sentimentalist metaethics. Then we will speak about how this allows for

sentimentalist reasons for action, both moral and self-regarding or self-interested. Those reasons will be no less valid as reasons than the metaethical moral realism (if any) that can be defended based on the notion of empathy.

Compassion and benevolence arguably work via empathy. As Hume in effect indicates in Book II of the *Treatise of Human Nature*, where we have stronger or more empathy our benevolent/compassionate concern is greater: as with our family and friends and people who are close to us in space and time. So as virtues (here we are into Book III) compassion and benevolence work via empathy. But these are what Hume calls natural virtues, and he contrasts them with what he calls artificial virtues like honesty/justice and fidelity to promises. However, we can reject that distinction in limning the moral virtues because Hume's account of the artificial virtues is very problematic (as even Hume himself to some extent recognized) and because (at least according to the book MS) a sentimentalist can reduce moral claims about justice, about respect, and even about deontology to facts/assumptions about empathy and empathically based concern for others. MS devotes a great deal of discussion to that project, and we don't intend to repeat that discussion here. The summarizing point, rather, is that empathy with others can be viewed as the basis of the moral virtue of moral agents. The next step is to argue that empathy also plays a metaethical role in moral thought and judgment. Hume made this assumption as the basis for his account of the semantics of moral utterances, and MS makes that assumption too, though in a somewhat different fashion from the way Hume does. It holds that moral judgment is grounded in second-order empathy with empathy or its absence. Hume points out at the end of the *Treatise* that we can be warmed by the warmth a friend may display in his actions toward his friend. The friend who displays the warmth in his actions arguably shows empathy with his friend's

needs or problems, and the observer then in some sense is empathically warmed by the warmth displayed by the helpful friend. This "empathy with empathy" is a kind of ur-approval. But Hume doesn't talk about the way we can be chilled by the lack of warmth or empathy some agent might display toward someone who needs their help. Being chilled by someone's cold-heartedness is also then second order and constitutes a kind of primitive form of disapproval.

From there we can get to moral realism by making adjusted use of Kripke's idea of reference fixing (Kripke 1970/1980). We cannot go into all the details—you can look at MS if you want more. But we will tell you what in overview can be said about using reference fixing to argue for a naturalistic sentimentalist moral realism. Ethical rationalists, going all the way back to Plato, often invoke an analogy between moral claims/knowledge and mathematical claims/knowledge. Empiricists and sentimentalists, however, often appeal to a proposed different analogy, that between color qualities/attributions and moral ones. We shall proceed along the latter lines through applying a good deal of what Kripke says about color terms to moral language. Kripke's reference fixing approach treats the colors of things as a completely objective matter rather than as involving a disposition to affect people in certain ways, and we shall argue that such an approach to moral terms, though not *completely* analogous with what Kripke says about color and "natural kinds," is analogous enough to allow us to ground a naturalistic variety of full-blown moral realism along sentimentalist lines.

Kripke argues that we rigidly fix the reference of a term like "red" as applied to things outside the mind via some description like "whatever it is in objects that typically causes red experiences (or perceiving redly) in us." If it turns out that a reflectance property *r* out there in objects is what tends to cause perceiving or experiencing redly, then that

external property is what (objective) redness is. (We here omit qualifications about the contexts of color perception.) And it is that property, it is redness, even in possible worlds where it causes green experiences in suitably altered humans or other possible animals, so objective redness is not relative to human experiential dispositions and is fully out there in objects.

Consider, then, what was said earlier, for example, about *ur*-approval as constituted by being empathically warmed by the warm actions of another person toward a third party. The second-order warmth felt by the observer is caused by first-order warmth in the agent, and simplifying matters by the assumption that no one is being deceived and perceptual apparatus is functioning veridically, we can say that the reference of “morally good” is fixed by the phrase “whatever it is in agents and their actions that causes *ur*-approval in observers.” Well, the only thing in agents or their actions that can cause *ur*-approval in observers is agential warmth or caringness. In particular, if something about an agent causes a feeling of warmth in an observer, that isn’t sufficient for saying that the observer *ur*-approves of the agent because it doesn’t entail that the warm feeling was obtained through empathy. If *ur*-approval can only be evoked via an empathic process (and everything is veridical), then the only thing in an agent that can possibly evoke *ur*-approval is warmth or warm-heartedness in the agent, and our argument to that effect is entirely *a priori*. It follows that if the reference of “morally good” is fixed in the manner suggested above, then we can know *a priori* that only agential warm caringness counts as moral goodness.

This is moral realism because the moral goodness thus specified doesn’t depend on anyone’s attitudes toward that goodness; it is objectively “out there” in the person judged morally good. And claims like “it is morally good to be concerned about people in distress and make efforts to help them” have the

desirable additional property of being *a priori* true or valid, just what ethical rationalists and perhaps thoughtful others want to say about moral claims. Kripke treats the claim that objective redness is reflectance factor/property *r* as necessary but *a posteriori*, but the way we have proposed for fixing the reference of moral claims (and what we have said about moral goodness could be said in analogous terms about moral right and wrong) allows such claims, in a philosophically desirable way, to count both as objectively valid and as *a priori*.

Moreover, the property identified with moral goodness is not only a completely natural one, but also one that the sentimentalist will want to treat as central or foundational to their account of moral virtue and moral rightness. Obviously, the notion of caringness has to be specified in a way we haven’t done here, but such specification, as Slote has elsewhere argued at considerable length, can be effected in an *a priori* fashion by reference to the concept of empathy. Actions are wrong (not morally all right) if and only if they show the agent as lacking in empathy, and, as we just indicated, what it is to be lacking or not lacking in empathy for others can be spelled out, intuitively but carefully, in an *a priori* fashion. This then leaves us with the final task of showing how normativity applies within such a sentimentalist picture of objective morality.

Clearly, moral considerations can be inherently motivating on the account that has been offered. That someone one sees is suffering will motivate the observer via their empathy with the suffering person, so such considerations are motivationally normative for sentimentalism. They won’t be motivating, of course, for a psychopath who is totally lacking in empathy, but that needn’t make us qualify what was just said about inherent motivation, if the fact that psychopaths aren’t responsive to (non-natural) reasons doesn’t require ethical rationalists to qualify their

claims about inherent normativity. No one really knows what to say about psychopaths in relation to morality (e.g., do they possess moral concepts?), and both naturalists and non-naturalists can hold that moral considerations are motivationally normative if they can be said to automatically motivate any “normal” person.

Moral sentimentalism of the kind defended here can also account for the inherent motivational force of moral judgments or beliefs. These rest, we have argued, on second-order empathy, but second-order empathy cannot exist in the absence of first-order empathy. How could one be empathically warmed and moved by the empathic warm-heartedness of some agent unless one knows from one’s own case what it is to be empathically moved by the plight of some sufferer? So moral judgments inherently involve some degree of motivation that aligns with the judgment: where the motivation to help aligns, for example, with the judgment that it would be good to help.

The biggest challenge regarding normativity for any sentimentalist account of morality concerns normativity with respect to reasons. Where sentimentalism allies itself with non-cognitivism or error theory, the idea of objective, real reasons for action seems a non-starter, but the moral realism defended here on the basis of reference fixing precisely defangs the suspicion that sentimentalism cannot provide for genuine practical reasons. But then one wants to know more about how real reasons can be brought under the aegis of the sentiments. In the *Treatise* Hume says in more than one place that there cannot be such a thing as a practical reason, but he is in fact inconsistent on this point and in other places he claims we can understand rationality or reasons in a reductive manner by equating them with calm passions. We have elsewhere explained at some length how such a sentimentalist reduction of self-regarding practical reasons might be effected, as when you have reason to leave your house when it is on fire (Slote 2019). But assuming such

reduction in self-regarding cases, how is one to understand moral reasons for action?

Well, let’s say someone has reason to leave their burning house but finds the only exit door to be hard to open. They begin to bang on and shove at the door from the inside and start yelling for help, and we can imagine an outside observer hearing and seeing all this and responding by trying to help the trapped person escape through that one door. The trapped person has reason, we have assumed, to try to escape through the door, but where is the moral reason the situation gives the observer to try to help the trapped person get out of the house? Well, on a sentimentalist theory that stresses empathy, that reason is not difficult to find. The observer empathizes with the trapped person’s fear and anxiety; the loud banging and screams for help from inside the house help this occur. But if the trapped person’s emotions can convey themselves to the outside observer via empathy, why should their reason to want to leave via the door not also convey itself via empathy to the observer—who then has reason of a moral kind to help the trapped individual escape through the one door. This is reductionistic vis-à-vis the moral reason involved, but it is reductionistic in precisely the way that a sentimentalist account of morality in naturalistic terms would need to rely on. The moral realism need in no way be compromised by the reductionism.

The sentimentalist form of naturalistic moral realism outlined and defended here essentially depends on the dual character of empathy: as both cognitively receptive to the world of other human/sentient beings and as motivationally active with respect to that world. (If this reminds you of yin and yang, you are absolutely on the right track, but that is a story to be told in other venues.) The example of empathy also gives the lie to J. L. Mackie’s well-known objections to what he calls “objective prescriptivity,” the idea that knowledge of realities can be inherently motivating (Mackie 1977). With regard to the

case of empathy we have *shown* how knowledge can entail motivation, and that means that pace Mackie there can be no objection in principle to the possibility or reality of objective prescriptivity. Mackie's objection to what Plato says about our responses to the Form of the Good works, if at all, because of problems with Plato's metaphysics and not, as Mackie assumes, because Plato's view allows for objective prescriptivity. All in all then, it seems we can naturalize ethics if we make metaethical use of reference fixing and make both metaethical and virtue-theoretical use of the concept of empathy. If the above is on the right track, we can "have it all," can have all three of naturalism, realism, and normativity within a moral sentimentalist framework.

#### CONCLUSION

The question here arises where Cornell realism goes wrong and how exactly this error is circumvented given a naturalistic moral realism of the kind offered by Slote in *Moral Sentimentalism*. A key difference between Boyd's and Slote's naturalistic realisms is that Boyd seeks to extend the original causal theory defended by Kripke (1970/1980) and Putnam (1975) to moral terms, whereas Slote makes an important modification to the causal theory. Within the original causal theory, scientific claims like "water contains oxygen molecules" or "whales are mammals" are a posteriori necessities, which means that it is conceptually possible but not metaphysically possible that these claims are false. It was exactly the embrace of moral principles as a posteriori necessities that allowed the Cornell realists to respond to Moore's open question argument and reject moral eliminativism.

On Slote's view, moral claims like "It's wrong to torture babies" or "Only empathic concern counts as moral goodness" are not a posteriori necessities but rather a priori necessities, akin to scientific claims such as "Fundamental particles have mass," "Electrons have a negative charge," or "Perceptual experiences are conscious states."

An important implication of Slote's sentimentalism is that when conducting ethical inquiry, we do not (necessarily) seek to mimic the aims and methodology of the natural (or empirical) social sciences. While the empirical sciences cannot avoid relying on a priori methods sooner or later, the scientific method is first and foremost empirical, not a priori.

Slote's sentimentalism thus parts ways with Cornell realism in terms of which type of naturalism they are most eager to preserve. Cornell realists wish to hold onto a methodological naturalism that leaves room for moral realism. Within philosophy, methodological naturalism is a view about philosophical practice within some philosophical area A (Devitt 1994; Papineau 2009; Brogaard 2016; Loewer 2017). It states that A and science are pursuing the same ends and should use similar methods to reach those ends. If the aim of science is to increase our collective reservoir of a posteriori knowledge by testing theories against the empirical observations, then according to methodological naturalism, this is also the aim of A.

Unfortunately, what the Cornell realists wish to do is a lost cause because methodological naturalism cannot be reconciled with moral realism. This is because determining why an action is wrong or an agent evil needs to be done through careful reflection, not empirical investigation or archival research to dig up historical documents.

But not much is lost by rejecting methodological naturalism. Upon reflection it should become clear that the non-naturalism most of us feel appalled by is ontological non-naturalism. This is the kind of non-naturalism that allows for moral properties that do not globally supervene on descriptive properties and perhaps even allows for supernatural ethical entities that are causally efficacious. But no such appalling non-naturalism follows from a naturalistic moral realism along the lines of Slote's sentimentalism.<sup>15</sup>

*University of Miami*

## APPENDIX

Antirealism entails the existence of a properly placed agent in every possible world. As human existence is contingent, this necessary agent must be supernatural or divine (Compare Brogaard and Salerno 2005).

*Notation*

$p$ : Placeholder for sentence letters

$Q$ : Ideal circumstances that include a properly placed agent  $S$  obtain.

$Bp$ :  $p$  is rationally believed by  $S$ .

$\leftrightarrow$ : The strict biconditional (i.e., the necessary material biconditional).

$\boxtimes \rightarrow$ : The David Lewis conditional.

$\diamond$  and  $\boxtimes$  are dual operators. The accessibility relation  $R$  must be reflexive and transitive, but it need not be symmetric.

*Proof:*

- |     |  |   |
|-----|--|---|
| 1.  | $\forall p (p \leftrightarrow (Q \boxtimes \rightarrow B(p)))$ | Anti-Realism about Truth                  |
| 2.  | $(Q \leftrightarrow (Q \boxtimes \rightarrow B(Q)))$           | From 1 by substitution.                   |
| 3.  | $(Q \leftrightarrow (Q \boxtimes \rightarrow B(Q)))$           | From 2, given the reflexivity of $R$      |
| 4.  | $Q$  | Assumption for $\rightarrow$ Introduction |
| 5.  | $Q \boxtimes \rightarrow B(Q)$                                 | From 3, 4, by detachment                  |
| 6.  | $B(Q)$   | From 4, 5 by detachment                   |
| 7.  | $Q \rightarrow B(Q)$   | From 4, 6, by $\rightarrow$ Introduction  |
| 8.  | $\boxtimes(Q \rightarrow B(Q))$                                | From 2, 3–7, by closure                   |
| 9.  | $Q \boxtimes \rightarrow B(Q)$                                 | From 8                                    |
| 10. | $Q$  | From 2, 9, by attachment                  |
| 11. | $\boxtimes \boxtimes(Q \rightarrow B(Q))$                      | From 8, given the transitivity of $R$     |
| 12. | $\boxtimes(Q \boxtimes \rightarrow B(Q))$                      | From 8–9 11, from closure                 |
| 13. | $\boxtimes(Q)$   | From 2, 3, & 10–11 12, by closure         |

## NOTES

- Others sympathetic to Cornell realism include Railton (1986) and Thomson (1997; 2001; 2008).
- See Shafer-Landau (2003).
- The term is due to Levine (1983). See also Chalmers (1995).
- For discussion of the abundant view of properties, see McGrath (2006). Thanks to a reviewer for encouraging us to emphasize that while Cornell realism is a form of (indirect) reductionism, it is not a form of eliminativism (van Riel et al. 2019). Cornell realism is thus consistent with a materialistic worldview. In fact, as Boyd himself points out, if this worldview is true, then “in some sense all natural phenomena are “reducible” to basic physical phenomena” (Boyd 1988: 194).
- The kind of naturalism defended by Cornell realists is the metaethical cousin of known as “a posteriori materialism” in contemporary philosophy of mind. A posteriori materialism, or what David Chalmers (2003) calls “type-B materialism,” has been defended by (among many others) Levine (1983), Loar (1990/1997), Papineau (1993), Tye (1995), Lycan (1996), Hill (1997), Block and Stalnaker (1999), and Perry (2001). Phenomenal truths, on the latter view, are said to strongly supervene on microphysical truths without being derivable from the latter, even in ideal inferential circumstances (Chalmers 2003; Chalmers 2009). A posteriori materialism is commonly classified as a form of strong naturalism (with weak naturalism being the view that only physical properties can be causally efficacious). For discussion see also Devitt, (1994), Papineau (2009), Brogaard (2016), Loewer (2017)

6. Moore (1903, 5–21).
7. Thanks to an anonymous reviewer here.
8. Alvin Plantinga (1969) was the first following Ruth Barcan Marcus (1961; 1963) to argue that *ordinary* proper names are rigid designators, as argued by Quentin Smith (1998, 274). Smith's paper stirred up heated debate and protests when first presented on Dec 28 1994 at an Eastern Division Meeting of the American Philosophical Association. Marcus had previously but without any apparent success made similar points in a published letter.
9. Variations on this sort of argument can be found in Hare (1952), Dreier (1990), Smith (1994), Brogaard (2007).
10. See also Dreier (1990) and Unger (1995). For discussion see Brogaard (2007; 2012; 2017).
11. One way to block this proof is to reject Lewis' assumption that counterfactuals with impossible antecedents are vacuously true. On an alternative view, counterfactuals with impossible antecedents are to be treated as sometimes true, sometimes false. On this characterization, it is no longer straightforward to prove that  $(Q \Leftrightarrow (Q \boxtimes \rightarrow B(Q)))$  entails  $\boxtimes Q$ .
12. Williams (1985), Dancy (1995), Väyrynen (2019).
13. See e.g., Bloom (2016).
14. See e.g., Arendt (1968), Viroli (1995), Bloom (2016). Arendt regards both emotional empathy and compassion as excuses to *evade* action.
15. We are grateful to two anonymous reviewers and Mirco Sambrotta for comments on an earlier version of this paper.

## REFERENCES

- Almeder, R. 1990. "On Naturalizing Epistemology," *American Philosophical Quarterly*, vol. 2, no. 4, pp. 263–279.
- Antony, L. 1993. "Quine As Feminist: The Radical Import of Naturalized Epistemology," in *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, ed. Louise Antony & Charlotte Witt, Boulder: Westview Press, pp. 185–226.
- Arendt, H. 1968. *Men in Dark Times*, New York: Harvest Books/Harcourt Brace.
- Bentham, J. 1789/1907. *An Introduction to the Principles of Morals and Legislation*, Oxford: Clarendon Press.
- Boyd, R. 1988. "How to Be a Moral Realist," in *Essays on Moral Realism*, ed. G. Sayre-McCord Ithaca, NY: Cornell University Press, pp. 181–228.
- Blackburn, S. 1984. *Spreading the Word*, Oxford: Clarendon.
- Block, N. and Stalnaker, R. 1999. "Conceptual analysis, dualism, and the explanatory gap." *Philosophical Review*, vol. 108, pp. 1–46.
- Bloom, P. 2016. *Against Empathy: The Case for Rational Compassion*, New York: Ecco.
- Brink, D. O. 1989. *Moral Realism and the Foundations of Ethics*, New York: Cambridge University Press.
- . 2001. "Realism, Naturalism, and Moral Semantics," *Social Philosophy and Policy*, vol. 18, pp. 154–176.
- . 2007. "The Autonomy of Ethics," in *The Cambridge Companion to Atheism*, ed. M. Martin, Cambridge: Cambridge University Press, pp. 149–165.
- Brogaard, B. 2007. "Moral Contextualism and Moral Relativism," *The Philosophical Quarterly*, vol. 58, no. 232, pp. 385–409.
- . 2012. "Moral Relativism and Moral Expressivism," *Southern Journal of Philosophy*, vol. 50, no. 4, pp. 538–556.

- . 2016. “Against Naturalism about Truth,” in *The Blackwell Companion to Naturalism*, ed. K. Clark, Oxford: Blackwell, pp. 262–276.
- . 2017. “Moral Contextualism and Epistemic Contextualism: Similarities and Differences”, in *Routledge Handbook of Epistemic Contextualism*, ed. J. J. Ichikawa, Routledge, pp. 361–374.
- Brogaard, B. and Salerno, J. 2005. “Anti-Realism, Theism and the Conditional Fallacy,” *Nous*, vol. 39, no. 1, pp. 123–139.
- Chalmers, D. J. 1995. “Facing Up to the Problem of Consciousness,” *Journal of Consciousness Studies*, vol. 2, no. 3, pp. 200–219.
- . 2003. “Consciousness and its Place in Nature,” in *The Blackwell Guide to the Philosophy of Mind*, ed. S. Stich and F. Warfield, Oxford: Blackwell.
- . 2009. “The Two-Dimensional Argument Against Materialism,” in *Oxford Handbook of the Philosophy of Mind*, ed. B. McLaughlin, A. Beckermann, and S. Walter, Oxford: Oxford University Press, pp. 313–335.
- Dancy, J. 1995. “In Defense of Thick Concepts,” *Midwest Studies in Philosophy*, vol. 20, no. 1, pp. 263–279.
- Devitt, M. 1994. “The Methodology of Naturalistic Semantics,” *Journal of Philosophy*, vol. 91, no. 10, pp. 545–572.
- Donnellan, K. S. 1966. “Reference and Definite Descriptions,” *Philosophical Review*, vol. 77, pp. 281–304.
- Dretske, F. 1995. *Naturalizing the Mind*, Cambridge, MA: MIT Press.
- Dreier, J. 1990. “Internalism and Speaker Relativism,” *Ethics*, vol. 101, pp. 6–26.
- Flikschuh, K. & Ypi, L. (eds.) (2015). *Kant and Colonialism: Historical and Critical Perspectives*, Oxford: Oxford University Press.
- Foot P. 2001. *Natural Goodness*, Oxford: Clarendon Press.
- Garrett, A. and Sebastian, S. 2017. “David Hume on Race,” in *The Oxford Handbook of Philosophy and Race*, ed. Naomi Zack, Oxford: Oxford University Press.
- Geach, P. 1969. “The Perils of Pauline,” *Review of Metaphysics*, vol. 23, no. 2, pp. 287–300.
- Green, T. H. 1883. *Prolegomena to Ethics*, ed., A. C. Bradley, Oxford: Clarendon.
- Hare, R. M. 1952. *The Language of Morals*, Oxford: Clarendon Press.
- Harman, G. “Moral Relativism Defended,” *Philosophical Review*, vol. 84, pp. 3–23.
- Hill, C. S. 1997. “Imaginability, Conceivability, Possibility, and the Mind-Body Problem,” *Philosophical Studies*, vol. 87, pp. 61–85.
- Horgan, T. and Timmons, M. 1991. “New Wave Moral Realism Meets Moral Twin Earth.” *Journal of Philosophical Research*, vol. 16, pp. 447–465.
- . 1992a. “Troubles on Moral Twin Earth: Moral Queerness Revived,” *Synthese*, vol. 92, no. 2, pp. 221–260.
- . 1992b. “Troubles for New Wave Moral Semantics: The ‘Open Question Argument’ Revived,” *Philosophical Papers*, vol. 21, no. 3, pp. 153–175.
- Hume, D. A. 1739–40. *A Treatise of Human Nature*, Oxford: Oxford University Press, 2000.
- . 1777. Essay XXI “Of National Characters,” in *Essays, Moral, Political and Literary*, London: A. Millar.
- Kaplan, D. 1989. “Demonstratives,” in ed. J. Almog, H. Wettstein, & J. Perry, *Themes from Kaplan*, New York: Oxford University Press, pp. 481–563.
- Kant, I. 1785/1997. *Groundwork of the Metaphysics of Morals*. Trans. M. Gregor. Cambridge: Cambridge University Press.
- . 1797/1996. *The Metaphysics of Morals*. Trans. M. Gregor. Cambridge: Cambridge University Press. 6:462.
- Kim, J. 1988. “What is Naturalized Epistemology?” *Philosophical Perspectives*, vol. 2, pp. 381–406.
- Kitcher, P. 1992. “The Naturalists Return,” *Philosophical Review*, vol. 101, no. 1, pp. 53–114.



- Kraemer, E. R. 1990. "On the Moral Twin Earth Challenge to New-Wave Moral Realism," *Journal of Philosophical Research*, vol. xvi, pp. 467–472.
- Kripke, S. 1970/1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press. *Naming and Necessity* is a transcript of three lectures, given by Kripke, at Princeton University in 1970.
- Laurence, S., Margolis, E. & Dawson, A. 1999. "Moral Realism and Twin Earth," *Facta Philosophica*, vol. 1, pp. 135–165.
- Levine, J. 1983. "Materialism and qualia: The explanatory gap," *Pacific Philosophical Quarterly*, vol. 64, pp. 354–61.
- Loar, B. 1990/1997. "Phenomenal states," *Philosophical Perspectives*, vol. 4, pp. 81–108. Revised edition in N. Block, O. Flanagan, and G. Güzeldere, (eds.), *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- Loewer, B. M. 2017. "A Guide to Naturalizing Semantics," In *A Companion to the Philosophy of Language*, ed. C. Wright and Bob Hale, Oxford: Blackwell, pp. 174–196.
- Lutz, M. and Lenman, J. 2018. "Moral Naturalism," in *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/fall2018/entries/naturalism-moral/>.
- Lycan, W. G. 1986. "Moral Facts and Moral Knowledge," *Southern Journal of Philosophy*, vol. 24, pp. 79–94.
- . 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*, New York: Penguin.
- Marcus, R. Barcan. 1961. "Modalities and Intensional Languages," *Synthese*, vol. 13, no. 4, pp. 303–322.
- . 1963. "Classes and Attributes in Extended Modal Systems," Proceedings of a Colloquium on Modal and Many-Valued Logic. *Acta Philosophica Fennica*, vol. 16, pp. 123–136.
- McGrath, M. 2006. "Propositions," in *The Stanford Encyclopedia of Philosophy* (Winter 2006 Edition), ed. Edward N. Zalta, URL = <http://plato.stanford.edu/archives/win2006/entries/propositions/>.
- McPherson, T. 2019. "Supervenience in Ethics," in *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/win2019/entries/supervenience-ethics/>.
- Moore, G. E. 1903. *Principia Ethica*, Cambridge: Cambridge University Press.
- Papineau, D. 1993. "Physicalism, consciousness, and the antipathetic fallacy," *Australasian Journal of Philosophy*, vol. 71, pp. 169–83.
- . 2009. "Naturalism," in *The Stanford Encyclopedia of Philosophy* (Spring 2009 Edition), ed. E. N. Zalta, URL = <http://plato.stanford.edu/archives/spr2009/entries/naturalism/>.
- Perry, J. 2001. *Knowledge, Possibility, and Consciousness*. Cambridge, MA: MIT Press.
- Plantinga, A. 1969. "De Re Et De Dicto," *Noûs*, vol. 3, no. 3, pp. 235–258.
- . 1982. "How to Be an Antirealist," *Proceedings and Addresses of the American Philosophical Association*, vol. 56, no. 1, pp. 47–70.
- Putnam, H. 1975. "The Meaning of 'Meaning'." In *Language, Mind and Knowledge*, ed. K. Gunderson, Minneapolis: Minnesota University Press, pp. 131–193.
- Quine, W.V.O. 1969. "Epistemology Naturalized," in W.V.O. Quine, *Ontological Relativity and Other Essays*, New York: Columbia University Press, pp. 69–90.
- Railton, P. 1986. "Moral Realism," *Philosophical Review*, vol. 95, no. 2, pp. 163–207.
- Rawls, J. 1971. *A Theory of Justice*, Cambridge, MA: Harvard University Press.
- Schiffer, S. 1987. *Remnants of Meaning*, Cambridge, MA: MIT/Bradford.
- Shafer-Landau, R. 2003. *Moral Realism: A Defence*. Oxford: Oxford University Press
- Slote, M. 2010. *Moral Sentimentalism*, Oxford: Oxford University Press.
- . 2014. *A Sentimentalist Theory of the Mind*, Oxford: Oxford University Press.
- . 2019. "Sentimentalist Practical Reason and Self-Sacrifice," *International Journal of Philosophical Studies*, vol. 26, no 3, pp. 419–436.

- Sonderholm, J. 2013. "Unreliable Intuitions: A New Reply to the Moral Twin-Earth Argument," *Theoria*, vol. 79, no. 1, pp. 76–88.
- Spencer, H. 1891. *Essays: Scientific, Political and Speculative*, 3 vols., London: Williams & Norgate.
- Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell.
- Smith, Q. 1998. "A More Comprehensive History of the New Theory of Reference," In *The New Theory of Reference: Kripke, Marcus, and its origins*, ed. P. W. Humphreys & J. H. Fetzer, Kluwer Academic Publishers, pp. 235–283.
- Street, S. 2006. "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Review*, vol. 127, pp. 109–166.
- . 2008. "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About," *Philosophical Issues*, vol. 18, pp. 207–228.
- Sturgeon, N. 1985. "Moral Explanations," in *Morality, Reason, and Truth*, ed. D. Copp and D. Zimmerman, Totowa, NJ: Rowman and Allanheld, pp. 49–78.
- Thomson, J.J. 1997. "The Right and the Good," *Journal of Philosophy*, vol. 94, pp. 273–298.
- . 2001. *Goodness and Advice*, Princeton, NJ: Princeton University Press.
- . 2008. *Normativity*. Chicago: Open Court.
- Timmons, M. 1999. *Morality without Foundations: A Defense of Ethical Contextualism*. New York: Oxford University Press.
- Tye, M. 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
- Unger, P. 1995. "Contextual Analysis in Ethics," *Philosophy and Phenomenological Research*, vol. 58, pp. 1–26.
- van Riel, Raphael and Van Gulick, Robert. 2019. "Scientific Reduction," in *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/spr2019/entries/scientific-reduction/>.
- Väyrynen, P. 2017. "The Supervenience Challenge to Non-Naturalism," in *The Routledge Handbook of Metaethics. Routledge Handbooks in Philosophy*, ed. T. McPherson, and D. Plunkett, Routledge, Abingdon, Oxon, UK, pp. 170–184.
- Väyrynen, P. 2019. "Thick Ethical Concepts," in *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), ed. E. N. Zalta, URL = <https://plato.stanford.edu/archives/sum2019/entries/thick-ethical-concepts/>.
- Viroli, M. 1995. *For Love of Country: An Essay on Patriotism and Nationalism*. Oxford: Oxford University Press.
- Williams, B. 1985. *Ethics and the Limits of Philosophy*. New York: Oxford University Press.