REALISM, METASEMANTICS, AND RISK

Billy Dunaway

OES REALISM about a subject matter entail that it is especially difficult to know anything about it? In broad outline the motivation for an affirmative answer to this question is a natural one: since realism (on a superficial gloss) holds that a domain exists independently of what we think, experience, or feel, it is possible for beliefs about the domain to diverge systematically from the facts. Realism about a subject matter, then, entails that the relevant facts are independent of us in a way that allows for widespread and systematic error in our beliefs about them.

This is just an initial gloss on a common and natural view about the epistemic consequences of realism. It is worth emphasizing at the outset that the alleged problem is a *substantive* connection between a kind of metaphysical thesis that is associated with the term "realism" and an epistemological claim according to which we lack—in a sense to be made more precise below—epistemic access to the facts, realistically construed. Thus I am conceiving of realism broadly as a metaphysical thesis. By way of contrast, some approaches to realism take the thesis to entail a lack of epistemic access *by definition*.[1] I am not working with these epistemic characterizations of realism here. Instead the question is whether the metaphysics of realism nontrivially implies that a range of beliefs are epistemically defective, and the aim is to investigate whether this argument constitutes a powerful epistemological consideration against the realist thesis.

"Realism" is a term that can apply (or fail to apply) to views about a wide variety of domains: physical objects, scientific unobservables, and mental states are all examples. The skeptical consequences of realism I sketched above allegedly follow from realism regardless of subject matter. I will focus, in what follows, on one particular domain where the skeptical consequences of realism are especially pressing and have received significant discussion. This is the domain of morality, and normativity more generally.[2] While the central characteristics of

---

1   See, e.g., Dummett, "Realism"; and Wright, *Truth and Objectivity*.

2   Examples of epistemological arguments against realism in this domain include Harman's claim that moral facts do not causally explain why we form moral judgments (Moral Expla-

normative language that give rise to epistemic difficulties for realism about normativity may also be present given realism about other domains, I will not raise this question here. It is, however, worth keeping in mind the question of whether the skeptical argument developed here applies to realism about other domains.

The argument I will develop has at its center the notion of *metasemantic risk*: this is the idea that our normative terms could have, in suitably different conditions, referred to something besides the normative properties they actually pick out.[3] I will outline how a particular kind of metasemantic risk follows from some core commitments of any plausible version of normative realism. And this kind of risk has consequences for knowledge and epistemic justification. I spell out these connections in an argument I call the *Argument from Risk*, and I will use the argument to explore the epistemology of realism.

1. RISK: METASEMANTIC AND EPISTEMIC

The short version of an argument that connects metasemantic risk and the absence of knowledge uses two technical terms, which I will describe briefly here, before adding more detail when evaluating the argument. *Metasemantic risk* refers to the possibility of a shift in the reference of a term. So while the normative term "ought" actually refers to obligation, there is some metasemantic risk in "ought" because the term could have referred to something else. That is, if "ought" is metasemantically risky, there is a possible linguistic community that speaks a language that is similar to English but differs from actual English enough that their "ought" refers to something distinct from obligation. Exactly how significant the risk is—that is, how similar this possible community where "ought" shifts reference is to our own—is a question I discuss below.

The second technical term is *epistemic risk*. A belief is subject to epistemic risk when it is at risk of being false in a way that is incompatible with that belief being knowledge. If I believe that I ought to keep my promise, but the belief is at risk in the relevant (epistemic) sense, then I could have had a false belief about my

---

nations of Natural Facts"); Street's argument that it is compatible with a naturalistic evolutionary process that we make moral judgments according to any of a wide variety of mutually incompatible moral systems ("A Darwinian Dilemma for Realist Theories of Value"); and Mackie's "argument from queerness," which (in one form) claims that moral properties are massively different in kind from any other property we know about (*Ethics*). I will not address these arguments here, and instead aim to present a distinct epistemological worry for realism.

3 Hawthorne makes use of a related notion ("*A Priority* and Externalism"). While the context of Hawthorne's discussion is a slightly different one—his focus is on a characterization of *a priori* knowledge, not epistemological arguments against realism—the present paper owes much to Hawthorne's discussion.

promise-keeping obligations. It follows, by definition, that I do not know that I ought to keep my promise.

The Argument from Risk connects metasemantic risk with epistemic risk for normative beliefs. According to the argument, metasemantic risk for normative terms is a commitment of realism about normativity. Epistemic risk in normative beliefs is an alleged consequence of metasemantic risk. In worlds where "ought" has shifted reference, some agents will have false normative beliefs, and this makes normative beliefs suffer from epistemic risk. So metasemantic risk implies that normative beliefs are not knowledge.

1.1. The Argument and an Illustration

The main premises in this argument are as follows:

1. "Ought" is metasemantically risky.
2. If "ought" is metasemantically risky, then one could easily be in a world where "ought" does not refer to obligation.
3. If one could easily be in a world where "ought" does not refer to obligation, then one could easily have had a false normative belief.
4. If one could easily have had a false normative belief, then one's actual normative beliefs are at epistemic risk and are not knowledge.

In the rest of this paper, I will spell out why the Argument from Risk is not a straightforward instance of a general argument that can be applied without modification to any domain. Instead, its premises are in a number of ways very plausible when their subject is *normative* belief, because of some unique features of our normative thought and language. Further, some of the premises in the Argument from Risk will be much more plausible to someone who adopts a *realist* view about the metaphysics of normativity: in fact, I will argue that realists must accept some of the premises in order to successfully respond to other arguments against realism in the literature. None of this is to deny that metasemantic risk may give rise to epistemological worries in other domains as well. But the rationale behind such worries will not necessarily be analogous to the support I offer for the premises in the Argument from Risk here.

Before turning to an evaluation of the Argument from Risk, we can begin with a concrete case where metasemantic risk appears to give rise to epistemic risk. (The case is loaded with theoretical assumptions that I will discuss later; the purpose here is only to provide an intuitive illustration of the epistemological problems that arise if the assumptions are correct.) Suppose that our community uses "ought" to refer to obligation, and moreover that among the obligatory actions is the act of giving 10 percent of one's annual income to charity. But giving 25 per-

cent of one's income to charity is not, we can suppose, obligatory. Because uses of "ought" are subject to metasemantic risk, there are worlds where our use of "ought" changes slightly. Among these worlds there are some where the linguistic situation is such that our use of "ought" refers to a property distinct from obligation—call it *obligation**—that has giving away 25 percent of one's income in its extension.

Our term "ought" refers to obligation. For reasons I will discuss below, it is plausible that "ought" also refers to obligation in many other possible worlds that we could easily have been in, which differ only slightly from our world with respect to how we use our word "ought." But the presence of metasemantic risk means that in some worlds, "ought" refers to a distinct property. Perhaps in these worlds it refers to a property that applies to acts that would be best, without regard to whether these are acts that an agent can realistically perform. In such worlds speakers insist that "ought" applies not only to acts that a speaker can reasonably be expected to perform but also to acts that, regardless of physical limitations of actual agents, would be best if they were to occur. Such acts are obligatory*. The claim that "ought" is metasemantically risky would be witnessed by a possible community that manages to use their normative "ought" to refer to obligation*.

Giving 25 percent of one's income to charity is, while not obligatory, obligatory*. In a world where "ought" refers to obligation*, one speaks falsely if one says "giving 25 percent of one's income is not something one ought to do." One also *believes* something false if one forms the belief in the proposition this sentence expresses in the world in which "ought" has shifted reference. Since we could be in such a world, if the Argument from Risk is sound, the normative belief that one ought to give 10 percent of one's income to charity is subject to knowledge-destroying epistemic risk.

1.2. A Program for Filling in the Argument from Risk

Every term in our language is capable of referring to something other than what it refers to in English, since there is no intrinsic connection between a string of letters or phonemes and the reference-determining features of the term. We could have used "ought" as we actually use "cat"; if we did, "ought" would not be a normative term in our language. This is not (an interesting form of) metasemantic risk. The Argument from Risk requires the possibility of semantic shifts that could easily have happened, and which, if they did happen, would give rise to false *normative* beliefs.

There are independently plausible theses about how we use normative language, and how we form our normative beliefs, that make the semantic shifts in normative language more interesting than a generic case of change in reference. In broad outline, one distinctive feature of normative language is that it

is possible to use normative language in many different ways, without making a conceptual mistake: there are, for example, possible communities of speakers that coherently apply their normative "ought" to acts of selfishly keeping one's money for oneself. This is an extreme difference between us and other coherent users of normative language, but there are all kinds of differences in between: some apply their "ought" to acts of keeping one's money when the benefit to oneself is extremely large; others make slightly less demanding exemptions, and so on. Each of these communities might still be motivated to do what they say they "ought" to do in the right way, and each community is not conceptually confused, so their "ought" will still have the role of a normative term. I will say that possible uses of normative language are *modally continuous*.

It is natural to add a second claim to this, which I will argue below is a commitment of any defensible version of normative realism on independent grounds. While uses of normative language can differ in all kinds of ways, many possible communities that use their "ought" as a normative term still manage to refer to *obligation*, rather than some distinct property. The community that says "one ought to be selfish and not donate any money to charity" manages to say that not donating has the property of *being obligatory*, the same property that we refer to when we say "one ought to give 10 percent of one's income to charity." They say something false about the same thing we are talking about, rather than saying something potentially true about a distinct property that fits their use better. I will say that normative terms are *semantically stable*.

However, there are limits to the amount of stability in any term, and normative terms are no exception. Thus I will say that normative terms are *moderately* semantically stable, since there are some possible communities that use their "ought" in ways that make it refer to something distinct from obligation. This is a significant assumption, and I return to it below.

Third, we and other possible linguistic communities can rely on the meaning of our public language term "ought" to form beliefs about what it refers to—whether this is obligation or some other property. That is, when one is in a community whose "ought" refers to obligation and accepts the sentence "one ought to give to the poor," one will typically have the corresponding normative belief that giving to the poor is obligatory. If one were to be in a part of a possible community whose normative "ought" refers to obligation*, then in accepting the sentence "one ought to give to the poor," one would typically have the belief that giving to the poor is obligatory*.

Just as the Argument from Risk needs a refined notion of metasemantic risk, it also needs refinements to the notion of *epistemic* risk. The generic notion of *risk* concerns what happens in nearby worlds, or worlds that could easily have ob-

tained.[4] Risking a false belief is, subject to refinements, incompatible with knowledge. Similarly, one way to lose justification for a belief is to learn that it could easily have been false in a sense that is incompatible with knowledge. So metasemantic risk will plausibly have important epistemological consequences for both normative knowledge and normative justification. But epistemic risk is not simply a matter of having a false belief in a nearby world. Once we add the needed refinements, the ways in which normative beliefs are metasemantically risky will make the Argument from Risk more compelling in the case of normative belief specifically.

2. PREMISE 1: METASEMANTIC RISK

Premise 1 in the argument from risk says:

    1. "Ought" is metasemantically risky.

Metasemantic risk takes the following form for normative terms: they are stable, but the stability is only moderate. In this section I will sketch the motivations for both parts of this premise.

    There are several considerations that suggest that stability is an explanatory *desideratum* for a realist view.[5] I will focus on the Moral Twin Earth case, but recent literature adds further considerations in favor of stability.

2.1. Moral Twin Earth

In a series of papers including "Troubles on Moral Twin Earth" and "Troubles for New Wave Moral Semantics," Horgan and Timmons argue that certain versions of realism cannot explain the range of disagreement between possible communities that use moral language. They argue against a version of realism that includes a causal theory of reference, due to Boyd, by describing two possible communities whose use of moral vocabulary is causally related to different properties but who nonetheless appear to disagree about morality:

    Earthlings' moral judgments and moral statements are causally regulated

---

4   Cf. "safety" principles in Sosa, "How to Defeat Opposition to Moore"; Williamson, *Knowledge and Its Limits*; and Pritchard, *Epistemic Luck*.

5   Note that even if stability is an explanatory *desideratum*, it does not follow that every realist view in fact explains it. Some views may acknowledge stability as an explanatory goal and treat it as a cost if they fail to explain its full range. Railton, "Moral Realism," is an example of a realist view that acknowledges the limits to the range of stability it predicts for moral terms. That a view has some theoretical costs is not, on its own, a decisive reason to reject it; Enoch emphasizes this methodological point (*Taking Morality Seriously*). However, I will assume that the realist does not have to concede stability as a point in favor of competing views.

by some unique family of functional properties, whose essence is functionally characterizable via the generalizations of a single substantive moral theory. Suppose, too, that this theory is discoverable through moral inquiry employing coherentist methodology. For specificity, let this be some sort of consequentialist theory, which we will designate T^c.

Now for Moral Twin Earth. Its inhabitants have a vocabulary that works very much like human moral vocabulary; they use the terms "good" and "bad," "right" and "wrong," to evaluate actions, persons, institutions, and so forth (at least those who speak twin English use these terms, whereas those who speak some other twin language use terms orthographically identical to the corresponding moral terms in the corresponding Earthly language). But on Moral Twin Earth, people's uses of twin-moral terms are causally regulated by certain natural properties distinct from those that (as we are already supposing) regulate English moral discourse. The properties tracked by twin English moral terms are also functional properties, whose essence is functionally characterizable by means of a normative moral theory. But these are *non-consequentialist* moral properties, whose functional essence is captured by some specific deontological theory; call this theory T^d.[6]

Horgan and Timmons think that when we consider communities like these, it is clear that they disagree about morality: "Here the question about what really is the fundamental right-making property seems to be an open question, and one over which Earthlings and Twin Earthlings disagree."[7]

There are a few details that are needed to turn the Moral Twin Earth case into an argument for stability for normative terms. First, although Horgan and Timmons are explicitly concerned with moral terms like "good," similar points apply to normative vocabulary like the all-things-considered "ought."[8] Second, the intuition of disagreement is not limited to the single case involving the Earthlings and Moral Twin Earthlings. As Horgan and Timmons emphasize elsewhere, similar cases can be described involving other pairs of possible communities whose use of moral vocabulary differs in other ways, aside from being causally regulated by different properties.[9] Finally, a realist should want to explain the

---

6   Boyd, "How to Be a Moral Realist"; Horgan and Timmons, "Troubles on Moral Twin Earth," 245.

7   Horgan and Timmons, "Troubles on Moral Twin Earth," 248.

8   Cf. Dunaway and McPherson, "Reference Magnetism as a Solution to the Moral Twin Earth Problem."

9   Horgan and Timmons, "Copping Out on Moral Twin Earth." How far the disagreements extend is an interesting question. I will address this point below.

disagreement by explaining how it is that each community in a Moral Twin Earth scenario is referring to the same property, and thereby makes a claim that is incompatible with claims made by speakers in the other community.

These details together motivate a realist view that entails a measure of semantic stability for normative terms. Since, for a realist, these disagreements should be explained in part by a metasemantic theory that entails that each community is referring to the same property, it follows from an adequate realist treatment of Moral Twin Earth cases that "ought" is semantically stable.[10]

Similar lessons emerge from more recent discussions of normative objectivity and knowledge, which I will mention only briefly. One comes from what I will call a "symmetry argument," found in Eklund's *Choosing Normative Concepts*. Eklund's "Bad Guy" is a possible user of a normative "ought" who, like the Twin Earthlings in the Moral Twin Earth scenario, uses the term differently from actual users. Bad Guy ends up saying different things than we say, applying "ought" for example to acts of stealing from the poor. Moreover, Bad Guy acts as we would expect for someone who applies a normative term to such acts.

If Bad Guy were speaking *truly*, by referring to a property that is distinct from obligation, there would be a kind of symmetry between him and us, since each of us speaks truly by using our normative "ought" and acts accordingly. This, according to Eklund, should be troubling for the realist: Bad Guy does things that he ought not to do. There should be some grounds for criticizing him. But, as Eklund points out, Bad Guy can make symmetrical criticisms of us: we fail to do some things that are obligatory*, and Bad Guy, in his language, speaks truly when he says "they fail to do some things that are obligatory."[11]

10 This is a commitment that is specific to realism: a noncognitivist or expressivist might explain the disagreements differently, cf. the notion of "disagreement in plan" in Gibbard, *Thinking How to Live*.

11 Here is Eklund:

> We can still say that Bad Guy doesn't do what he all-things-considered ought to do or has reason to do. But using his language, Bad Guy can say the corresponding things about us. Using his counterpart of "wrong"—the word in his vocabulary that has the role for him that "wrong" has for us—he can say that we do "wrong" things. And he is as correct in his verdict about us as we are in our verdict about him. The same would go for all other normative vocabulary.... Despite all the realist trappings that our normative language is supposed to have, there may still for all that has been said be *parity* between us and Bad Guy that the ardent realist would want to avoid. For all that has been said, Bad Guy is not objectively mistaken about anything; he just does not employ our notion of reason or our notion of what ought to be done but instead employs alternative normative notions. (*Choosing Normative Concepts*, 5)

Clarke-Doane, *Morality and Mathematics*, explores additional worries along similar lines.

I have not argued here that a realist can meet these explanatory *desiderata*, and I will, in what follows, simply assume that they can be met.[12] The Argument from Risk claims that it follows that realism faces a further problem: by adequately accounting for stability, realism introduces an epistemological difficulty—namely, the inability to know normative claims. Thus the Argument from Risk captures a distinct (alleged) problem for realism since it *assumes* that the realist can explain the moderate stability of normative language that is raised by Horgan and Timmons, and claims that it therefore fails to explain how normative knowledge is possible.[13]

2.2. Interlude: Realism and Stability

Stability is an explanatory *desideratum* for the realist. Nonrealists, such as expressivists who follow Gibbard, can also endorse stability for normative terms.[14] In this case the endorsement will be explained with nonrealist resources: for example, in Gibbard's terms, the explanation will involve the thesis that claims about reference are *plan-laden*.[15] So simply explaining stability does not make a view realist; whatever additional metaphysical claims are needed for realism will need to be compatible with stability for normative terms. The Gibbardian explanation in terms of normative beliefs as planning states will not do for the realist.

For the sake of illustration, here is one version of a theory that can explain the stability datum for normative terms for realists. The theory consists in the metaphysical claim that some properties are metaphysically privileged or *elite* properties and the metasemantic claim that elite properties are easier to refer to

12 Elsewhere I argue that the realist has the resources to meet this challenge, together with other explanatory *desiderata* I describe below. See Dunaway, *Reality and Morality*.

13 Eklund does raise an epistemological issue that emerges from his discussion of Bad Guy. The issue is not what he calls "run-of-the-mill" skepticism concerning knowledge of what we ought to do (*Choosing Normative Concepts*, 14). Instead, Eklund's "normative skepticism" is a special question, since he grants that we can know what we ought to do but is less sanguine about our ability to know, roughly, that we should care about what we ought to do (as opposed to, say, caring about what Bad Guy is talking about with his normative terms). This is just a gesture at what Eklund's epistemological issue is, since Eklund thinks that the real question may be "ineffable" and so knowing its answer, which settles what is at issue between us and Bad Guy, will be difficult if not impossible (25).

This is an interesting question, but it is not the one I will address in what follows: I am concerned with "run-of-the-mill" skepticism, which is our inability to know what we ought to do. Moreover, I will not be presupposing any of Eklund's discussion of the existence of a further, ineffable issue that is under dispute between us and Bad Guy, as Eklund does. The epistemological problems I raise here can be raised even if we reject the existence of such an issue.

14 Gibbard, *Meaning and Normativity*.

15 See Dunaway, "Expressivism and Normative Metaphysics," for elaboration.

than nonelite properties. These claims together are sometimes labeled *reference magnetism* and are outlined by David Lewis.[16] This is the view that terms refer to what they do not solely in virtue of how they are used but also in virtue of how the world is: some speakers' use of a term best fits a particular property *P*, but there is an elite property *P**, similar to *P*, that fits their use pretty well and is metaphysically privileged over *P*. These speakers are, according to a metasemantic theory that includes reference magnetism, referring to *P**. This is the sense in which elite properties are easy to refer to: we can refer to elite properties without tailoring our usage to fit them precisely.

If the normative realist holds that obligation is a metaphysically elite property, then obligation (modulo the additional assumption that there are no additional elite normative properties in the vicinity of it) will be easy to refer to for speakers that use normative language. Views along these lines have been sketched by Van Roojen, Edwards, and Dunaway and McPherson.[17] I will not rehash the details here, but the theory illustrates one substantive set of claims that appears to explain stability for the realist. It is, plausibly, not an unmeetable explanatory *desideratum*.

2.3. Moderate Stability

The normative term "ought," I will assume, is metasemantically risky because the stability is only *moderate*. "Ought," as used by some possible communities, refers to some property other than obligation. But what makes the stability merely moderate is the fact that some of these possible communities that use their "ought" to refer to something other than obligation still use it as a *normative* term. Roughly, this means that they use the term to settle what to do and to close off deliberation. When one uses "ought" in this way, it is not coherent to conclude that one "ought" to act in a certain way and yet fail to do so.[18] This is to use "ought" with a *normative role*.

Some additional notes about this assumption are in order.

First, the standard motivations for stability do not motivate a thesis that is stronger than moderate stability. The standard examples of possible communities that disagree in Moral Twin Earth scenarios all involve communities that appear to accept different substantive theories about what wrongness, or obligation, consists in. (The usual example involves consequentialists and deontolo-

---

16 Lewis, "New Work for a Theory of Universals" and "Putnam's Paradox."

17 Van Roojen, "Knowing Enough to Disagree"; Edwards, "The Eligibility of Ethical Naturalism"; and Dunaway and McPherson, "Reference Magnetism as a Solution to the Moral Twin Earth Problem."

18 Cf. Gibbard, *Meaning and Normativity*.

gists, but examples can be multiplied by imagining that the communities accept different first-order ethical theories.) However, there are other ways for possible communities to differ while still using their "ought" with a normative role. Generalization off the usual cases will not give us any reason to think that all of these possible communities will be referring to obligation.

As a second point, we can give an example of this kind of community that appears to (i) use their "ought" with a normative role and (ii) refer to something distinct from obligation. Suppose that, as a matter of fact, our "ought" does not apply to single-handedly ending a large famine. There are a number of possible communities that do apply their "ought" to ending the famine. Some of these communities fit the model of the Horgan and Timmons Moral Twin Earth case, where the community simply has a different substantive theory of obligation. But there are other possibilities: the community could value the same things we value, but not use their normative "ought" as if it were governed by the principle "ought" implies "can." If the difference between English speakers and speakers in this possible community is that the latter say "one ought to end the famine single-handedly," and this difference arises because they do not restrict application of their "ought" to actions that can be performed, then the difference in how they use their "ought" need not be a difference over substantive first-order theory. The difference between them and actual English speakers does not constitute a dispute over what makes an action obligatory.[19]

All of this is compatible with both communities—including the community that does not have an "ought" that is governed by the "ought" implies "can" principle—using their "ought" with a normative role.[20] Cases like the community in the previous example suggest that it is not at all clear that a community should be interpreted as referring to obligation simply because they use "ought" with a normative role. English speakers can agree that options they are unable to perform have a property that resembles the property that makes actions obligatory. It is not inconceivable that some other possible communities have normative terms that refer to such normatively significant properties that are distinct from obligation.

A final point is that in order to deny that "ought" is only moderately stable, one has to adopt an extreme metasemantic view. Such views exist in the literature: Wedgwood and Williams hold a version of the "conceptual role determinism for wrongness" thesis.[21] Likewise, Eklund takes such a thesis seriously

---

19 See Dunaway, *Reality and Morality*, chs. 1–2, for more discussion.

20 They will conclude that they "ought" to do some things that they inevitably fail to do. But this is not incompatible with the term having a normative role in their language; they are simply forced to conclude that they regularly display a kind of incoherence.

21 Wedgwood, "Conceptual Role Semantics for Moral Terms" and *The Nature of Normativity*;

without outright endorsing it.[22] I will not argue against these views here, but it is worth noting that they require a highly unusual metasemantic contribution from a term's normative role. This is because on each view, the fact that "ought" is used with a normative role is sufficient for "ought" to refer to obligation; all other facts about how the term is used are irrelevant to what it refers to.[23]

Perhaps normative language is special in this way. But theories that entail that normative role determines reference are more ambitious than what the intuitive data supports, and they treat normative role as a privileged reference-determining feature of normative terms; I will assume without further argument here that the stability of normative terms does not extend as far as these theories imply.[24] However, for readers who are not inclined to grant this assumption, the Argument from Risk promises to provide additional motivation for such a strong metasemantic assumption. If the argument is successful, it shows that there are *epistemic* considerations that favor a strong metasemantic commitment to stability, in addition to the usual arguments that motivate such views.

3. PREMISE 2: RISK OF SHIFTED REFERENCE

The first step in the Argument from Risk is to show that metasemantic risk for normative terms exists. The second step is to claim that a reference shift could easily have happened. This is what premise 2 claims:

> 2. If "ought" is metasemantically risky, then one could easily be in a world where "ought" does not refer to obligation.

It is worth emphasizing that while reference shifts are possible for other nonnor-

---

Williams, "Normative Reference Magnets" and *The Metaphysics of Representation*. Quote from Williams, "Normative Reference Magnets," section 2.3.

22 Eklund, *Choosing Normative Concepts*.

23 For example, with color terms, reference is partly determined by something like conceptual role: it is part of the conceptual role of "red" that it refers to a color that is darker than what "orange" refers to and lighter than what "purple" refers to. But this alone does not settle that "red" refers to redness—cf. the "permutation problem" in Smith, *The Moral Problem*. Additional facts about how "red" is used by English speakers, including the fact that they regularly apply their "red" to red objects, are relevant.

24 Note that conceptual determinism for wrongness (or obligation) is extreme from a realist perspective, when obligation and other normative properties are real properties that serve as candidate referents for a term or concept. In this case there are alternative properties that are candidate referents; explaining why a term used with a normative role never refers to these alternative candidate referents, regardless of the additional features of its use, is a substantial task. Eklund calls such a view "metasemantically radical" (*Choosing Normative Concepts*, 167).

mative portions of our vocabulary, these shifts are much less threatening from an epistemological perspective.

Reference shifts for color terms like "red" are, in a sense, much more widespread than shifts for "ought." Any community-wide shift in usage is guaranteed to produce a different referent; color terms are not stable, but rather are semantically *plastic*.[25] If we had applied "red" to a slightly different range of light wavelengths (and made corresponding adjustments in our use of other color terms), then we would have referred to something other than redness. The analogue of semantic stability for "red"—that we would have continued to refer to redness and would be making false claims about which objects are red—is not plausible.

These shifts do not threaten our knowledge of colors. In most cases I will not be prevented from knowing that my coffee mug is red on the grounds that my term "red" could easily have referred to a different property.[26] One reason is that these are *known* shifts. Since we know that whenever a community-wide change in use occurs, the reference of "red" changes as well, worlds where "red" refers to some property besides redness are also worlds where we are in a position to update our belief-forming practices to reflect the change in subject matter. For instance, in worlds where we apply "red" to some slightly orangey shades, we refer to a different property—call it red*. But we will also reliably believe, of those orangey shades, that they have the property red*, since we are aware that reference shifts with this change in use. We will not systematically have false beliefs where semantically plastic terms are involved.[27]

25 I am borrowing this terminology from Dorr and Hawthorne, "Semantic Plasticity and Speech Reports."

26 See Clarke-Doane, *Morality and Mathematics*, ch. 6, for a claim along similar lines that rejecting stability in favor of a plenitudinous ontology of set-like entities solves epistemological problems for set-beliefs. Clarke-Doane explicitly considers the prospects for an analogous plenitudinous ontology of obligation-like properties. But these advantages are available only to someone who explicitly rejects stability for "ought," a consequence I am assuming the realist wants to avoid.

27 Another reason is that competent users of color terms will allow for a wide range of borderline cases. For shades on the borderline between red and orange, a competent user of "red" will refuse to apply "red" to these shades and will refuse to apply "not red" to them. Likewise, she will not believe of these borderline shades that they are red. If "red" in her community shifts reference to a slightly different property, even if she is not aware of the shift, she will not have a false color belief. Rather, even though her term refers to the color redness* rather than redness, redness* will include some borderline cases of redness about which she withholds belief—she will not falsely believe that these shades are red*. It would take an extremely large semantic shift for "red" to refer to a property that includes shades that, in the actual world, she believes are not red.

Our color beliefs are, of course, not infallible. We can be mistaken about what the term "red" in our own language refers to. But there is no systematic risk here, as the errors will be

Semantic shifts for stable normative terms appear to be much more threatening epistemologically. When we form beliefs about what we ought to do, having a good belief-forming disposition requires being resistant to changing one's beliefs whenever one's community uses normative terminology slightly differently. One should not be disposed to change one's beliefs about what one ought to do just because one's community uses "ought" slightly differently. In general, being disposed to change normative beliefs to match the way one's community uses "ought" is a way to have *false* beliefs about obligation, since our obligations do not for the most part depend on our linguistic habits.

While these semantic shifts do occur, it is unlikely that one will know the precise point at which this happens. The difference between stable and plastic terms does not concern whether we can know how our community uses them. Rather, the difference lies in whether we will be able to tell that the reference of these terms has shifted for a community that uses their terms differently. With a semantically stable term like "ought," we cannot simply assume that any unanimous shift in use in our community is accompanied by a shift in reference. As a result the disposition to (correctly) treat "ought" as semantically stable will, in some worlds where the term has shifted reference, yield a false positive: one will, in virtue of having the disposition, continue to treat "ought" as referring to obligation.

To sum up, premise 2 in the Argument from Risk is true, in the following sense: there are nearby worlds where "ought" refers to something distinct from obligation because these worlds are not easily distinguishable from worlds where it refers to obligation. These are worlds where one will continue to use "ought" as if it refers to obligation, by speaking as if there are no meaningful differences between the usage of "ought" in one's community and the usage of the term in a community that refers to obligation. The differences between such worlds are small but significant: *ex hypothesi*, the usage of "ought" in one's own community, when a semantic shift has occurred, makes it the case that "ought" does not refer to obligation.

4. PREMISE 3: FROM METASEMANTIC RISK TO RISK OF A FALSE BELIEF

Premise 3 in the Argument from Risk says:

3. If one could easily be in a world where "ought" does not refer to obligation, then one could easily have had a false normative belief.

This is a conditional claim with an antecedent that concerns normative *language*,

---

attributable to speakers having impoverished or misleading information, or failing to treat their color terms as sufficiently plastic. I return to the significance of this in sections 4 and 5.

and a consequent that is about the propositional attitude of *belief* about normative matters. According to this premise, shifts in the reference of a piece of language have consequences for the content of our beliefs. The connection is not obvious, so in this section I will outline a background picture of the connection between language and belief that makes premise 3 plausible.

The previous section outlines how metasemantic risk gives rise to worlds with false normative beliefs because, given the moderate semantic stability of normative vocabulary, we could have been in a world where "ought" refers to something besides obligation, but we do not know that this shift has occurred. As a schematic example, take a world where the term "ought" shifts reference by referring not to obligation, but rather to obligation*.[28] Premise 3 claims that the content of normative beliefs of language users in a community that uses "ought" to refer to obligation* will change. Believers will usually rely on the referents of terms in their public language in order to form beliefs about the world around them. And so in worlds where the term "ought" shifts reference to refer to obligation*, speakers in these worlds will be forming normative beliefs about obligation* as well.

A shift in the referent of "ought" in a public language does not entail that one has the same linguistic dispositions as all other members of the community one is a part of. In a world where usage of "ought" by a community has shifted enough to make the term refer to obligation*, members of that community will form normative beliefs about obligation* by relying on the referent of "ought" in their language. For instance, suppose Suzy forms a belief about the normative status of giving 25 percent of her income to charity. This is the belief that has the content she would express by uttering the sentence "it is not the case that one ought to give 25 percent of one's income to charity." If Suzy is in a world where the referent of "ought" has shifted, the normative belief she forms in that world does not have the content *giving 25 percent of one's income to charity is not obligatory*. Instead she refers to—and forms beliefs about—something else, namely obligation*. Shifts in the usage of normative language by Suzy's community can affect shifts in what Suzy's normative thought is *about*, without Suzy changing how she forms her own beliefs. The shift can be a result of changes in her surrounding linguistic community only.

Moreover, these shifts can affect whether Suzy's normative beliefs are *true*. In a world where "ought" has not shifted reference, and refers to obligation, Suzy's

---

28 Earlier I suggested that if the normative "ought" refers to the property of producing the most good out of the options an agent has a reasonable ability to perform, the referent of "ought" could shift in some worlds to refer to the property of producing the most good out of all possible actions, whether performable or not. The schematic example can be filled in by identifying this property with obligation*.

normative belief has the content of her sentence "it is not the case that one ought to give 25 percent of one's income to charity." She forms a true belief. The belief she forms in the world where her community uses "ought" slightly differently, and thereby refers to obligation*, need not be true. In such a world, Suzy need not be aware that in this world she refers to something different from what her counterpart in an obligation-referring world refers to, and she will continue to believe the content of the sentence "it is not the case that one ought to give 25 percent of one's income to charity." Thus she might continue to form beliefs in the same way. But in the shifted world, "ought" refers to obligation*. Giving 25 percent of one's income to charity *is* obligatory*. Suzy then has a false belief, and the only difference between this world where she has the false belief and a world where she has a true belief is that the usage of normative vocabulary by others in her linguistic community has changed.

As I have emphasized, reference shifts are possible for nonnormative vocabulary as well, but there are some distinguishing features that make the possibility of false beliefs owing to these nonnormative reference shifts less threatening. Return to the example of color terms, like "red," as a contrast case. These terms are much more widespread than shifts for "ought," as almost any community-wide shift in usage is guaranteed to produce a different referent. If we had applied "red" to a slightly different range of light wavelengths (and made corresponding adjustments in our use of other color terms), then we would have referred to something other than redness. This is what makes color terms semantically plastic.

The consequences of plasticity for color terms give rise to a difference in status of color *beliefs* in worlds where reference has shifted. In particular, the increased chance of a reference shift in color terms has the opposite effect: speakers will be less likely to form false beliefs when there are reference-shifting changes in their community's usage. The semantic shifts in plastic terms are *known* shifts, and so we know (roughly) that whenever a community-wide change in use occurs, the reference of "red" changes as well. Moreover, since one should not be disposed to change one's beliefs about what one ought to do in the possible situation where one's community uses "ought" slightly differently, someone with the right dispositions will be further susceptible to false normative beliefs when shifts in the reference of "ought" do occur. These points suggest that when semantic shifts for "ought" do occur, we will be at risk not only of using "ought" accordingly but also of having false normative beliefs.[29]

29 Of course in other nonnormative areas, the relevant vocabulary might not be as plastic as color terms, and instead will display some degree of stability. Steadfast dispositions might be appropriate to some degree as well. These are questions that will need to be answered in extending the Argument from Risk to nonnormative areas.

5. PREMISE 4: FALSE BELIEFS AND EPISTEMIC RISK

The final step in the Argument from Risk claims not only that moderate stability gives rise to possible false normative beliefs but also that the presence of these false beliefs is incompatible with *knowledge*. The risk of false belief is an *epistemic* risk, which prevents even those normative beliefs that manage to be true from being knowledge. This is captured by premise 4:

> 4. If one could easily have had a false normative belief, then one's actual normative beliefs are at epistemic risk and are not knowledge.

The relevant notion of epistemic risk is captured broadly by a "safety" condition on knowledge, as developed in Sosa, Williamson, and Pritchard: beliefs are subject to epistemic risk when they fail to be safe.[30] My aim here is not to defend the safety condition. Instead, I aim simply to show that plausible refinements on such a condition, which are motivated by plausible intuitions about knowledge, are compatible with the claim that metasemantic risk gives rise to epistemic risk. For those who do not wish to think in terms of a connection between knowledge and safety, the motivating examples are still relevant, as they will provide constraints on alternative views of what is required for knowledge.

Below I will sketch two refinements that any plausible version of a safety condition will need to incorporate. These are a restriction of nearby false beliefs—the kind that make actual beliefs subject to distinctively epistemic risk—to beliefs that are (i) similar in content and (ii) the products of similar causal processes. Normative beliefs are plausibly subject to epistemic risks owing to semantic shifts, even when these refinements are in place. It is not obvious that the same goes for nonnormative beliefs.

The first qualification on the notion of epistemic risk is that a belief is at risk only if *similar* beliefs are false in nearby worlds.[31] I can know that I had breakfast

---

30 Sosa, "How to Defeat Opposition to Moore"; Williamson, *Knowledge and Its Limits*; Pritchard, *Epistemic Luck.*

31 What makes a world *nearby* in the sense relevant to epistemic risk? Williamson, who accepts a general connection between knowledge and the absence of risk, says that the concept of knowledge cannot be *defined* in terms of the absence of false belief in nearby (or "sufficiently similar") worlds:

> If one believes p truly in a case a, one must avoid false belief in other cases sufficiently similar to a in order to count as reliable enough to know p in a. The vagueness in "sufficiently similar" matches the vagueness in "reliable" and "know." . . . We need not assume that we can specify the relevant degree and kind of similarity without using the concept *knows*. (*Knowledge and Its Limits*, 100)

While I am not pursuing the question of whether knowledge can be analyzed in terms of

this morning even if there is a nearby world where I misremember the name of a new acquaintance and have a false belief about her name.

While misremembering an acquaintance's name does not put my beliefs about my breakfast at risk, this is not simply because the beliefs are different. Knowledge-destroying false beliefs do not need to be identical in content. If one is guessing at the answer to questions about the sums of moderately large numbers, then one's correct guesses will not be such that there are nearby worlds where the same belief is false. If one correctly guesses that $634 + 399 = 1,033$, then one has a true belief, and moreover this very belief is not false in any nearby world (in all nearby worlds, $634 + 399 = 1,033$). So one cannot have a false belief that $634 + 399 = 1,033$ in any nearby world. Correctly guessing does not, however, bring knowledge. If one is guessing at the relevant sums, then even if one actually gets the answer right, there is a nearby world where one instead comes to believe a related but false claim—for instance, that $634 + 399 = 893$. This belief is quite similar to one's actual belief. Since there are nearby worlds where one has false beliefs like this when one is guessing, one's actual true belief that $634 + 399 = 1,033$ is subject to risk.

Return to the case of normative beliefs that are subject to metasemantic risk. We can call a world in which one has false normative beliefs owing to semantic shifts in the normative "ought" a *shifty world*. The content of a normative belief in a shifty world is not identical to the content of our actual normative beliefs. Normative beliefs in the actual world are about obligation, but beliefs in the shifty worlds are about a different property.

The false beliefs in shifty worlds will, however, be similar enough to put our actual normative beliefs at risk. If our normative beliefs in the actual world are about obligation, shifty worlds where we instead form normative beliefs about the distinct property obligation* are still worlds where our normative beliefs are very similar: both beliefs involve reference to properties that broadly bear on what to do, or what is best. Moreover, these are formed with the use of concepts that bear on motivation to act, and so are plausibly still normative beliefs rather than beliefs about a radically different subject matter. So if beliefs about obligation* in the shifty world are false, they will be beliefs that are not disqualified

epistemic risk here, it is worth emphasizing that Williamson's line is one that is consistent with the approach to the central questions of this paper. That is, in asking whether normative beliefs can be knowledge for the realist, I will make claims about certain normative beliefs being false in nearby worlds. Williamson may be right that these judgments rest (partly) on our judgments about whether those that hold the relevant beliefs *know*. It is not my aim here to settle this question; rather the point is to deploy the framework of epistemic risk, which has been developed elsewhere, to draw attention to some connections between the metasemantic aims of the realist and normative knowledge.

from being beliefs that prevent our actual normative beliefs from being knowledge simply in virtue of being beliefs with distinct contents.

Here is a second qualification: not all similar false beliefs in nearby worlds are incompatible with knowledge. Some nearby similar false beliefs are arrived at in a suitably different way, and so do not put one's actual beliefs at risk in the relevant sense. If I happen to see a friend who usually lives in another city walk by, I know they are in town (we can suppose this is true even if I have no other evidence that they are visiting). This belief is not at risk just because there is a nearby world where our paths never cross and I continue to believe that my friend is not in town.[32] The reason is that the beliefs are formed by very different processes. The causal processes leading up to the formation of beliefs about my friend's location are very dissimilar, as one involves perception and the other involves an inference on the basis of my knowledge of my friend's usual place of residence. It is thus very natural to conclude that if a belief is subject to epistemic risk, there must be nearby false beliefs that are both similar in content *and* similar in respect to the causal processes that produce them.

Recall the difference between true normative beliefs and their counterparts in shifty worlds that are false owing to metasemantic risk: the only difference between them need be that the latter are formed in a world where the surrounding community of language users deploys normative vocabulary differently. Community-wide changes like this need not be accompanied by a change in one's own belief-forming methods; one can continue to use "ought" and form normative beliefs as if this term referred in one's language to obligation.

Moreover, there is no relevant difference in the aptness of the process one uses in the world where "ought" has shifted reference. Because steadfast dispositions are appropriate for normative terms and beliefs, there is no obvious sense in which the false normative belief is formed by a defective process. Instead, the process by which one forms normative beliefs in each world involves dispositions one should have in each world, since one should not change one's normative beliefs simply in response to changes in one's community's use of normative language.[33] Thus, having a false belief owing to a semantic shift is compatible

---

32 Cf. Pritchard, *Epistemic Luck.*

33 An anonymous referee raises the possibility that metasemantic risk raises analogous skeptical worries for beliefs about natural kinds, such as arthritis. This may be—I am not taking a stand on the extent of the epistemic consequences of metasemantic risk here—but the distinctive relevance of steadfast dispositions to normative belief formation should not be overlooked. In a world where speakers use "arthritis" to refer to a different natural kind, arthritis*, one might form false beliefs about arthritis* by using the relevant concepts as if they refer to arthritis. If one does this while knowing roughly how one's community (including the relevant experts) uses "arthritis," one will be making a mistake by taking the change in

with nearly everything in the process that produces the false belief being identical to the belief-forming process in a world where the shift has not happened.

Take for example Suzy's false belief about the nonobligatoriness of giving 25 percent of one's income to charity. The token process that produces this belief—which includes the practical reasoning Suzy employs to arrive at her belief as well as other psychological factors that can influence normative belief formation—can be nearly identical in a shifty world, where "ought" refers to obligation*. The difference in content, and the difference in the truth conditions for these beliefs, depends only on factors that lie outside the causal chains that produce these beliefs. We need only to change how the surrounding community uses their normative vocabulary in order to describe a world where Suzy's normative beliefs refer to obligation* and are therefore false.

These points together suggest that we can mount a strong case that possible normative beliefs that are false owing to metasemantic risk entail that our actual normative beliefs are subject to epistemic risk.

6. LIMITS TO THE ARGUMENT FROM RISK

So far I have assembled the credentials for metasemantic risk as the key notion in an epistemological argument against normative realism. Since there are independent constraints that entail that normative terms should, on the realist view, be moderately stable, metasemantic risk exists on the realist view. This yields possible false beliefs that have certain distinguishing features of epistemic risks for our actual normative beliefs. These false beliefs are the products of changes in community usage that are not distant possibilities; they are similar in content to our actual normative beliefs, and they are the products of very similar causal processes. These considerations ward off any general strategies for dismissing the Argument from Risk.

The Argument from Risk does not, however, establish sweeping skepticism for the normative realist. Rather, its ambitions will have to be scaled back: while some normative beliefs do fail to be knowledge owing to metasemantic risk, the skeptical consequences of realism do not extend to every normative belief. Whether this restricted skepticism is problematic for realism, and vindicates traditional epistemological worries about the view, is a question I will revisit in the conclusion.

---

usage to be irrelevant to the reference of "arthritis." There is a significant difference between the nondefective method one might use to form beliefs about arthritis in a normal world and the defective beliefs one would use to form beliefs about arthritis* in a world where "arthritis" has shifted reference. But there is no analogous difference in the methods one uses in forming normative beliefs in worlds where "ought" shifts reference.

6.1. Contingent Epistemic Risk

Premise 2 in the Argument from Risk assumes that the core semantic features of normative language for the realist—moderate semantic stability and modal continuity—imply that "ought" could easily have shifted referents.

A shifty world is a world where our normative beliefs are false owing to a semantic shift in the referent of "ought." One could *easily* have been in a shifty world, in the following sense: differences between our world and a shifty world are potentially very small and are not necessarily changes that we can tell will constitute a shift in the reference of "ought." Moderate semantic stability guarantees the existence of the shifty world; our dispositions to treat "ought" as stable make us likely to treat "ought" as referring to obligation even in some shifty worlds. In one sense the shifty world is one we could easily have been in, since in a shifty world one might not know that one is in a shifty world.

If we are in a nonshifty world, we can say that the world we are in is *stable*. Premise 4 claims that shifty worlds are close enough to stable worlds, and so the normative beliefs we hold in stable worlds are subject to epistemic risk. But it is not sufficient for distinctively *epistemic* risk that one not be able to tell the difference between the stable and shifty worlds. The kind of epistemic risk that is incompatible with knowledge requires that the causal process that produces a belief must be similar to one that gives rise to false beliefs in worlds that are metaphysically close to the actual world. Guessing gives rise to epistemic risk, since one's guess could easily have produced a different belief. Even though hallucination is possible, relying on visual impressions to form beliefs does not give risk to epistemic risk: a world where one hallucinates need not be metaphysically close to the actual world.

Nothing in premise 2 of the Argument from Risk guarantees that all stable worlds are close to shifty worlds in the sense that is relevant to epistemic risk. It is consistent with our world being one where "ought" refers to obligation that all of the nearby worlds where usage of normative terms differs only slightly from ours are also stable. All premise 2 tells us is that if we were in a (possibly distant) shifty world, we would not necessarily know that we are in a shifty world. In this case, the risk of false belief brought about by shifty worlds is not threatening to knowledge, since the shifty worlds are modally distant and cannot generate the kind of epistemic risk that prevents actual normative beliefs from being knowledge.[34]

There are, however, some stable worlds that are metaphysically close to

34 Williamson emphasizes a general version of the point that simply because we would not know that we are in the shifty world if we were in one, it does not follow that the world is a close one that threatens our knowledge in the actual world (*Knowledge and Its Limits*, sec. 8.2). If we were to assume that our inability to discriminate the shifty worlds where we have

shifty worlds. In such worlds, the motivations for premise 2 will be relevant to our normative knowledge in these stable worlds. Premise 2 could be filled out appropriately so it is true in these stable worlds: that is, worlds that are both difficult for us to distinguish from shifty worlds and also modally close to shifty worlds. With appropriate modifications to the other premises in the Argument from Risk, we will have a version of the argument that is *contingently* sound. Its premises will not be true in all worlds but will be true in some.

Nonetheless this reveals a distinctive epistemic commitment in virtue of the core features of the realist view. Owing to metasemantic risk, we will need to admit that it is a contingent possibility that we lack normative knowledge. If we are in a world where a shifty world is nearby, then we will lack normative knowledge. The presence of knowledge-destroying metasemantic risk is an epistemic limitation distinctive to realism but not a necessary feature of the view.

6.2. Higher-Order Knowledge

There is a related skeptical conclusion that the realist must accept for every stable world, including the actual one. Call the worlds that could be reached by just small changes in use from our actual world the worlds that are in the *immediate vicinity* of ours. If there is a shifty world in the immediate vicinity of ours, we lack *first-order* normative knowledge. But as the previous subsection argues, there is nothing in the commitments of the realist that entails that it is necessary that there should be a shifty world in the immediate vicinity of a world where normative beliefs are formed. What is necessary is merely that there is a shifty world somewhere in modal space.

The existence of a shifty world somewhere in modal space does not by itself threaten our first-order normative knowledge. But it does threaten our higher-order normative beliefs. Suppose the shifty world is not in the immediate vicinity of the actual world but is in the immediate vicinity of a world in our immediate vicinity. We can then know that we ought to give 10 percent of our income to charity (this is first-order knowledge). We do not, however, *know* that we know that we ought to give 10 percent of our income to charity. The higher-order belief is at risk, epistemically. There are worlds in our immediate vicinity where it is false—these are the worlds that have the shifty world in their immediate vicinity. (In a world with a shifty world in its immediate vicinity, we do not know that we ought to give 10 percent of our income to charity. And in worlds with a world where we do not know that we ought to give 10 percent of our income to charity in their immediate vicinity, we do not know that we know that we ought to give

_____

false beliefs guarantees their closeness, skepticism would follow for reasons having nothing to do with moderate stability.

10 percent of our income to charity.) So, even if we are not at immediate risk of forming false normative beliefs, the realist will have to concede that some of our higher-order normative beliefs will fail to be knowledge.[35]

Whether the loss of merely higher-order normative knowledge is damaging to the realist is an open question. It does, however, point to one respect in which a realist view that entails the moderate semantic stability of normative terms will have a distinctive (and limited) epistemic profile.

6.3. Imprecise Knowledge

So far we have seen that the Argument from Risk is limited in what it establishes, in certain respects. It is only contingently sound at best when it targets our first-order normative beliefs. All we can be certain of is that, owing to metasemantic risk, we will lack some higher-order normative knowledge.

In addition, the Argument from Risk is not equally threatening to all of our normative beliefs. Even if we are (as a contingent matter) in a world that is very close to a shifty world, there are kinds of normative belief that will have a claim to survive as normative knowledge. Begin with the beliefs that are most threatened by the Argument from Risk. The example I have been working with is the belief that it is obligatory to give 10 percent of one's income to charity. This is just for illustrative purposes, and there are many beliefs like it that will be similarly threatened: for instance, the belief that one ought to avoid eating any kind of animal regardless of the effects for future animal suffering, or the belief that lying to avoid personal embarrassment is impermissible. Even if these beliefs are true, if a certain type of shifty world lurks nearby, they will not be knowledge.

But what they all have in common is that they are extremely *specific* beliefs. A specific belief is one that is incompatible with beliefs that make only slightly different cutoff points for obligatory action. For example, believing that one ought to give 10 percent of one's income to charity is incompatible with the belief that one is only obligated to give exactly 9 percent of one's income to charity. (If the latter belief is true, one is not obligated to give 10 percent, though it would perhaps be commendable to do so.) Likewise, we could have believed that we are only obligated to avoid eating mammals, or that we are only obligated to refrain from lying when the consequences are more serious than mere personal embarrassment.

Specific beliefs like these go false in shifty worlds. There are ways for "ought" to shift its reference while still being used as a normative term that make the

---

35   We might have to go even further away from the actual world to find worlds where we do not know that we ought to give 10 percent of our income to charity, owing to the proximity of shifty worlds. Regardless, we will not know that we know that we . . . that we ought to give 10 percent of our income to charity for the same reasons; some iteration of knowledge will fail.

belief with the content (in the shifty world) of "one ought to give 10 percent of one's income to charity" false in that world. If giving 10 percent of one's income to charity is obligatory but not obligatory*, then acts of giving 10 percent do not have the property "ought" refers to in the shifty world. The same problem does not necessarily affect nonspecific beliefs. Take for instance the belief that one ought to give between 1 and 30 percent of one's income to charity (where this belief is equivalent to the disjunctive belief that one ought either to give exactly 1 percent, or one ought to give exactly 2 percent, and so on). This belief has a *margin for error* built in—unlike the specific belief that one ought to give 10 percent of one's income, it is compatible with a variety of specific facts about one's obligations. Some beliefs with sufficiently wide margins for error will not have false counterparts in nearby shifty worlds.

Even if one's belief that one ought to give 10 percent of one's income to charity has a false counterpart in a nearby shifty world, it does not follow that the belief that one ought to give between 1 and 30 percent of one's income to charity does as well. This is because of the built-in margin for error: even though what one believes when one forms a normative belief in the shifty world is different—one does not believe that one *ought* to give between 1 and 30 percent of one's income to charity, since one's term "ought" does not refer to obligation—one forms a true belief anyway. If a normative "ought" shifts reference to refer to obligation*, one still forms a true belief, since there is some percentage of one's income between 1 and 30 percent that is such that donating it is also obligatory*.

This points to a further respect in which the Argument from Risk is limited. Premise 3 does not distinguish between which normative beliefs go false in a shifty world. But it should, since some normative beliefs with sufficiently wide margins do not go false even when "ought" shifts reference. Even in a setting where a specific normative belief is at risk, and thereby cannot be knowledge, a belief with margins for error built in might, in the very same setting, be knowledge. So a suitable modified Argument from Risk will not threaten all normative knowledge equally.[36]

7. CONCLUDING REMARKS

There are, then, some limits to the amount of skepticism that the Argument from Risk entails. It does not entail a wholesale rejection of knowledge about the target domain, although it does point to distinctive ways in which our knowledge may be limited owing to a realist metaphysics that implies that normative terms

---

36 Thanks to Levi Spectre for discussion of this point.

are moderately semantically stable. Where do the limits leave realism's episte-
mological credentials?

From one perspective, it saves realism from the damning conclusion that re-
alism entails wholesale skepticism. The skeptical conclusion is obviously unpal-
atable, since if it followed from the core commitments of realism that all norma-
tive beliefs are not knowledge, we would have strong reason to look elsewhere
for a normative metaphysics that makes room for normative knowledge.

However, from another perspective, the Argument from Risk does not leave
realism on a par with its competitors when epistemological considerations are
in play. It is worth emphasizing the respects in which settling for knowledge only
for claims that have sufficiently wide margins for error is not entirely satisfactory.
At first glance it seems to be a small concession to hold that while we might not
know specific normative propositions—such as the claim that we ought to give
10 percent of our income to charity—we can know related propositions with
margins for error built in. But in fact our normative knowledge is connected in
a variety of important ways to other aspects of our cognitive and practical lives,
and our beliefs with wide margins for error cannot sustain all of these connec-
tions. I will close with two examples.

One example of the way this skepticism ramifies comes from the conse-
quences for knowledge of other normative claims that are believed in typical
ways. It is plausible that when we form beliefs by deducing them from other be-
liefs, the output of the deduction is known only if the premises from which it
was deduced were known. (This principle is sometimes known as *counter-clo-
sure*.) Typical agents who hold the wide-margin-for-error belief that they ought
to give between 1 and 30 percent of their income to charity often do so on the
basis of a deduction from a specific normative belief—say, the belief that one
ought to give 10 percent of one's income to charity. But if the specific beliefs
are not knowledge, then by counter-closure the wide-margin-for-error beliefs
that are deduced from them will not be knowledge either. So by conceding that
some normative claims are not known owing to metasemantic risk, we might
be forced to hold that other in-principle-knowable claims are not in fact known.

The other example is from the practical side: it is common to hold that an
agent's reasons, of a certain kind, can only be claims that she knows or has some
epistemic access to. For instance, if Bob does not know that the police are ar-
riving at the crime scene, it is infelicitous to say that Bob's reason for fleeing the
crime scene was that the police were arriving.[37] If Suzy gives 10 percent of her in-
come to charity out of her sense of moral duty, it is tempting to say that her rea-

37 Hyman, "How Knowledge Works." This is sometimes called an ascription of a *personal* rea-
son for action.

son for giving the money is that she is obligated to give 10 percent of her income to charity. But this cannot be her personal reason for acting, if she does not know specific normative propositions like this. And Suzy might not be thinking about a proposition with margins for error built in, much less take such a claim to be her reason—most people do not give 10 percent of their income merely on the basis of the fact that they are obligated to give between 1 and 30 percent of their income to charity. So even a lack of knowledge of some normative propositions can deprive agents of some important personal reasons for action.

These considerations are not decisive, but they point to some respects in which even a limited skepticism about normativity might be troubling for realism. How troubling the limited skepticism is remains to be seen, but I have sketched two reasons in closing for holding that the skepticism cannot be dismissed as wholly irrelevant simply because it is not wholesale skepticism. Proponents of realism can of course avoid these results by denying moderate semantic stability for normative terms. There are, however, strong motivations elsewhere in the literature for realists to accept moderate semantic stability. What (some version of) the Argument from Risk shows is that there may be no version of realism that is entirely problem free. Endorsing moderate semantic stability, even if a satisfactory metaphysics and metasemantics can be found to support it, faces its own problems in light of its epistemological consequences.[38]

University of Missouri–St Louis
dunawayw@umsl.edu

REFERENCES

Boyd, Richard N. "How to Be a Moral Realist." In *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord, 181–228. Ithaca, NY: Cornell University Press, 1988.

Clarke-Doane, Justin. *Morality and Mathematics*. Oxford: Oxford University Press, 2020.

Dorr, Cian, and John Hawthorne. "Semantic Plasticity and Speech Reports." *Philosophical Review* 123, no. 3 (July 2014): 281–338.

Dummett, Michael. "Realism." In *Truth and Other Enigmas*, 145–65. London: Duckworth, 1978.

Dunaway, Billy. "Expressivism and Normative Metaphysics." In *Oxford Studies in Metaethics*, vol. 11, edited by Russ Shafer-Landau, 241–64. Oxford: Oxford University Press, 2016.

———. *Reality and Morality*. Oxford: Oxford University Press, 2020.

Dunaway, Billy, and Tristam McPherson. "Reference Magnetism as a Solution to the Moral Twin Earth Problem." *Ergo* 3, no. 25 (2016): 639–79.

Edwards, Douglas. "The Eligibility of Ethical Naturalism." *Pacific Philosophical Quarterly* 94, no. 1 (March 2013): 1–18.

Eklund, Matti. *Choosing Normative Concepts*. Oxford: Oxford University Press, 2017.

Enoch, David. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press, 2011.

Gibbard, Allan. *Meaning and Normativity*. Oxford: Oxford University Press, 2013.

———. *Thinking How to Live*. Cambridge, MA: Harvard University Press, 2003.

Harman, Gilbert. "Moral Explanations of Natural Facts—Can Moral Claims Be Tested against Moral Reality?" *Southern Journal of Philosophy* 24, no. S1 (Spring 1986): 57–68.

Hawthorne, John. "*A Priori* and Externalism." In *Internalism and Externalism in Semantics and Epistemology*, edited by Sanford C. Goldberg, 201–18. Oxford: Oxford University Press, 2007.

Horgan, Terry, and Mark Timmons. "Copping Out on Moral Twin Earth." *Synthese* 124, no. 1–2 (July 2000): 139–52.

———. "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived." *Philosophical Papers* 21, no. 3 (1992): 153–75.

———. "Troubles on Moral Twin Earth: Moral Queerness Revived." *Synthese* 92, no. 2 (August 1992): 221–60.

Hyman, John. "How Knowledge Works." *Philosophical Quarterly* 49, no. 197 (October 1999): 433–51.

Lewis, David. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61, no. 4 (December 1983): 343–77.

———. "Putnam's Paradox." *Australasian Journal of Philosophy* 62, no. 3 (September 1984): 221–36.

Mackie, J. L. *Ethics: Inventing Right and Wrong*. London: Pelican Books, 1977.

Pritchard, Duncan. *Epistemic Luck*. Oxford: Oxford University Press, 2004.

Railton, Peter. "Moral Realism." *Philosophical Review* 95, no. 2 (April 1986): 163–207.

Smith, Michael. *The Moral Problem*. Oxford: Wiley-Blackwell, 1994.

Sosa, Ernest. "How to Defeat Opposition to Moore." *Philosophical Perspectives* 13, Epistemology (1999): 141–53.

Street, Sharon. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127, no. 1 (January 2006): 109–66.

Van Roojen, Mark. "Knowing Enough to Disagree: A New Response to the Moral Twin Earth Argument." In *Oxford Studies in Metaethics*, vol. 1, edited by Russ Shafer-Landau, 161–94. Oxford: Oxford University Press, 2006.

Wedgwood, Ralph. "Conceptual Role Semantics for Moral Terms." *Philosophical Review* 110, no. 1 (January 2001): 1–30.

———. *The Nature of Normativity*. Oxford: Oxford University Press, 2007.

Williams, J. Robert G. *The Metaphysics of Representation*. Oxford: Oxford University Press, 2020.

———. "Normative Reference Magnets." *Philosophical Review* 127, no. 1 (January 2018): 41–71.

Williamson, Timothy. *Knowledge and Its Limits*. Oxford: Oxford University Press, 2000.

Wright, Crispin. *Truth and Objectivity*. Cambridge, MA: Harvard University Press, 1992.