

Realism, Meta-semantics, and Risk

Billy Dunaway
University of Missouri–St Louis

Draft of 28th February 2017

Does *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. By espousing realism, we commit to the view that the relevant facts are independent of us and do nothing to guarantee that there is no systematic error in our beliefs.

This is just an initial gloss on a common and natural view about the epistemic consequences of realism.¹ Turning this natural thought into an epistemological argument against realism requires digging deeper into the alleged reasons for which realism can plausibly be claimed to entail that our beliefs are, or are likely to be, epistemically defective.

The literature on realism about morality, and normativity more generally, contains a range of arguments which promise to do exactly this job. My aim in this paper is to argue that these existing epistemological arguments against realism about normativity are unsatisfactory on methodological grounds, and then to use the lessons from these failures to develop a new epistemological argument against realism, and explore its consequences.

The existing epistemological arguments against realism either fail to target core features of realism, or do not connect realism to familiar epistemological properties like knowledge or epistemic justification. So to begin, we need to formulate a plausible and general account of the core features of realism. And we need to formulate a plausible necessary condition on these familiar epistemological notions. An epistemological argument against realism about normativity can succeed, then, only if it shows that it follows from the core feature of realism that our normative beliefs fail to satisfy the conditions on knowledge and justification.

I will outline one argument which has these features, but has not been explored in the existing literature. This argument has at its center the notion of *meta-semantic risk*: roughly, the idea that our normative vocabulary and thought could have, in suitably different conditions, referred to something besides the normative

¹Here it is worth emphasizing that I am conceiving of realism broadly as a metaphysical thesis, in a sense that will be made more precise below. Some approaches to realism take the thesis to entail lack of epistemic access *by definition*—see e.g. Dummett (1963) and Wright (1992).

properties they actually pick out. I will argue that the existence of meta-semantic risk follows from the core features of realism. And it plausibly has consequences for the absence of knowledge and epistemic justification. I spell these connections out in an argument I call the Argument from Risk.

The Argument from Risk, then, represents an attempt to show that it is a commitment of realism as such—and not just specific versions of realism that have been defended in the literature—that normative facts are difficult, or impossible to know. It is, in other words, an attempt to characterize the (lack of a) relationship between realism and knowledge that fits the natural thoughts I described above.

But in return for this generality, I will close by arguing that we lose the sweeping conclusions that many existing epistemological arguments against realism have hoped for. These arguments have tried to show that knowledge of *any* normative claims is *impossible*, if realism is true, and then taken this conclusion to decisively favor competitors to realism. But careful reflection on the Argument from Risk shows, in ways I will describe below, that we cannot reach conclusions quite this strong—and so realism is not threatened by outright damning skepticism. But neither can it take a cut-and-paste approach to epistemology, claiming that whatever explains our widespread knowledge in other domains can do the same for normative realism. Instead the phenomenon of meta-semantic risk means that realists will have to admit that our access to normative facts is limited in ways our access to facts elsewhere is not.

1 Preliminaries

I will begin by looking at existing epistemological arguments against realism about normativity. An *epistemological argument* is, for present purposes, an argument that aims to establish that realism is unattractive on the grounds that it entails that we have little to no epistemic access to the normative. (What *realism* is, on the other hand, is a question I will postpone for later.) Examples of epistemological arguments against realism include:

- Harman's (1986) claim that moral facts do not causally explain why we form moral judgments;
- Street's (2006) 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;
- Mackie's (1977) "argument from queerness" which (in one form) claims that moral properties are massively different in kind from any other property we know about.

For reasons of space, I will not focus on the details of each argument. Rather I will ask what they can show, even if their premises are true. An epistemological argument against normative realism, as I am understanding it, would show that

it follows from a core feature of the realist view that normative beliefs fail to constitute knowledge, or lack justification, or else are missing some other familiar epistemological properties.

But all of these arguments fail to have this structure. Mackie's argument rests on the premise that normativity, for the realist, is unlike any other domain we know about. Perhaps this is plausible for non-naturalist views of the normative in the style of Moore (1903), but clearly naturalistic versions of realism such as those found in Railton (1986) and Schroeder (2007) make this premise look unappealing. Arguably Harman's and Street's arguments are likewise limited in the same way. Street insists that her argument targets views according to which normativity is *mind-independent*² And Harman's premise that normative facts fail to exert causal influence on normative beliefs will look suspect to a naturalist who accepts that normative facts reduce to natural facts.

Moreover the epistemological side of these arguments, qua epistemological arguments against realism, is lacking in important respects. For instance, even if a set of facts doesn't exert causal influence on our beliefs, as Harman's argument claims, it doesn't immediately follow that these beliefs fail to be knowledge, lack justification, or fail to instantiate other epistemologically significant properties.³

This illustrates what I take to be the gold standard for an epistemological argument against realism: such an argument should proceed from premises about the central, core features of realism—and not simply claims that are true of some, but not all, versions of realism. Moreover it should conclude that realism entails, on the basis of these core features, that the relevant beliefs lack familiar epistemological properties: they are not knowledge, or cannot be justified.

In the next sections I will set up the key notions in an argument that has this structure. The first is an understanding of what the central features of realism about the normative are. And the second is a natural and tractable notion of a necessary condition on epistemological success—knowledge or epistemic justification. Then I will argue that there is a very natural connection between the metaphysics of realism and epistemic success. This bridge is formed by a phenomenon I call *meta-semantic risk*. Epistemological arguments against realism become much more promising, I will argue, when they focus on the phenomenon of meta-semantic risk.

2 Realism: the metaphysics of eliteness

If we are interested in assessing the potential epistemic problems for realism about the normative, we need to start with a fully general characterization of what realism is.

Begin with some examples: familiar versions of realism about the normative include the non-naturalism found in Moore (1903), and more recently defended

²Street (2006: 111)

³Clarke-Doane (2012)

by Shafer-Landau (2003) and Enoch (2011). Although I won't spend much space articulating what exactly I take to be the core of the non-naturalist view, it is uncontroversial that the view (in part) entails that the normative does not *reduce* to natural properties.

The absence of a reduction cannot be the core of realism. There are versions of *naturalistic* realism as well, which on a straightforward interpretation differ from non-naturalism in precisely this respect, while still retaining the core features of realism. Examples include: Railton (1986), Boyd (1988), and Schroeder (2007). Since these versions of realistic naturalism do not share the anti-reductive aspect of non-naturalism, we need to look elsewhere for the core features realism.

It is often claimed that the distinguishing feature of realism is not a denial of a reduction to the natural, but rather a denial of dependence on a certain part of the natural world, namely the *mental*.⁴ But this is plausibly not the most perspicuous way to characterize the core features of realism. Examples of normative realism suggest this: for Railton (1986), moral facts depend in part on facts about what an idealized agent would want. Similarly the role of desire-promotion in Schroeder (2007) and homeostatic property clusters that involve the satisfaction human—including psychological—needs in Boyd (1988) suggest that, since these views are realist views about normativity, realism does not necessarily require independence from the mental.

These are just some preliminary considerations concerning some common but ultimately unpromising proposals about the core features of realism.⁵ Instead of pursuing these problems further, I will outline what I take to be the best alternative. The alternative characterization of realism I have in mind focuses not on whether (or what) the normative depends on if it is to be real; rather it focuses on alternative ideology to characterize the metaphysical status of the real.

Broadly, all of these theories entail that normative properties are not very gerrymandered, making seemingly arbitrary distinctions between the things that instantiate them and those that don't. Rather these properties are, in the terminology of Lewis (1983, 1984), relatively *natural*, or *elite*. Examples of properties that lack naturalness in Lewis's sense include astrological properties such as *being a virgo*, or highly disjunctive properties such as *being either prime or located in Mongolia*. These properties are not instantiated by entities that have anything interesting in common with each other, and do not by themselves explain any substantive phenomena. Normative realism, by way of contrast, entails that normative properties interesting, explanatory, and unified: either they do this by locating the normative among the fundamental, irreducible parts of reality (non-naturalists), or by providing reduction bases that are themselves fairly unified and explanatory (naturalistic realists).

Metaphysically elite properties are, on the rough characterization I am work-

⁴See Brink (1984), Cameron (2008), Jenkins (2005), Pettit (1991), and Wright (1992) for discussion of this idea.

⁵For more discussion, see Dunaway (MS).

ing with here, distinguished by their explanatory role. One important aspect of this explanatory profile which I will focus here is the role they play in the explanation of the content of language and thought. More specifically: highly elite properties are, all things considered, *easy to refer to* in comparison with properties that are not very elite. For example: color-properties are not very elite. The property of being a surface that reflects light wavelengths between 620 and 750 nm—which we refer to with our term ‘red’—is not any more elite than the property of being a surface that reflects light wavelengths between 610 and 740 nm, or any other range of wavelengths. By contrast the fundamental physical quantities such as negative charge are highly elite: a property that has an extension that is similar to the extension of negative charge but differs by including some non-negatively-charged things will not be an elite physical property.

This shows up in thought and language in the following way. If we our use of ‘red’ were slightly different—say, we consistently applied the term to some slightly orange shades by applying it to surfaces that reflect light wavelengths between 610 and 740 nm—then ‘red’ would refer to a property that includes some orangey shades. (Moreover our thoughts formed by tokening the concept RED will have a different content that does not involve the property redness.) But if we were to use our term ‘negative charge’ slightly differently, physicists would not thereby be talking about a property distinct from negative charge. Slight changes in usage do not change the subject-matter of physics.

This phenomenon has been called *reference magnetism* in the literature, and I will use this terminology here.⁶ If realism is true and normative properties are highly elite, then they are reference magnets. Some communities that use normative language slightly differently will nonetheless refer to the same normative properties. This idea has significant import for traditional debates about reference and disagreement with normative terms.⁷

I take commitment to the eliteness of normative properties and the consequences of eliteness for normative thought and language to be core features of realism about the normative. This sets the stage for a fully general epistemological argument against realism that targets its core features, rather than ancillary claims that some realists accept, but others can deny.

At the core of the argument I will develop are two theses connected to the status of normative properties as reference magnets. This will then give us the resources to derive some consequences for normative epistemology, which I turn to in the next section.

2.1 *Moderate semantic stability*

The first thesis is that normative terms like ‘ought’ are *moderately semantically stable*. Begin with the semantic stability of ‘ought’: if the property ‘ought’ refers

⁶See Lewis (1983) for an initial presentation of this idea, and Hawthorne (2006, 2007) and Williams (2007) for discussion.

⁷See Dunaway and McPherson (2016) for elaboration and defense of this idea.

to is a reference magnet, then there are some possible communities of normative language users that use 'ought' differently than we do, but still refer to obligation. This is semantic stability: although there is a property that is distinct from obligation that fits their use better than obligation does, that property is not very elite. Thus these communities are talking about the same thing we are talking about, even though obligation is not the property that best fits how they use 'ought'.

But this stability is only *moderate*. While there are some changes in use of normative vocabulary that do not change what it is about, there are other possible worlds where use of 'ought' changes in ways which are sufficient to constitute a distinct reference for the term. Moreover there are possible changes in reference for communities to continue use 'ought' as a *normative* term—where these speakers are motivated to do the actions that it applies to, criticize others who fail to do such actions, and so on. When communities use 'ought' as a normative term sufficiently differently, so that there is an elite property distinct from obligation that better fits their use, they are no longer talking about the same thing we are.

Here is an example to illustrate this schematic characterization of the semantic profile of normative terms for the realist. Suppose that 'ought', in our mouths, refers to the property of maximizing happiness; only actions that produce the most happiness are in the extension of 'ought'. For the realist, this means that maximizing happiness is highly elite, and hence that it is a reference magnet. Thus there are possible communities like ours, which use 'ought' slightly differently—sometimes, we can suppose, they decline to apply 'ought' to an action that maximizes happiness. Nevertheless their term still refers to the maximization of happiness, since this property is more elite than the property that better fits this community's use of 'ought'. This is the semantic stability of 'ought'.

Nevertheless there are possible communities that continue to use 'ought' as a normative term, but are such that their use is better fit by a property distinct from happiness-maximization yet which is just as elite. For example take the property of being an action that maximizes happiness and is also *psychologically feasible* to perform. This is, we can suppose, just as elite as happiness-maximization simpliciter. This is the property of being an act that maximizes happiness, and moreover is one that humans with a typical psychology will usually be able to perform. If this property is just as elite as happiness-maximization, then it too is a reference magnet. There will be a possible community that uses 'ought' as a normative term, but their usage better fits the property of psychologically feasible happiness-maximization. Speakers in such a community don't refer to happiness-maximization with their use of 'ought'. Hence 'ought' is only moderately semantically stable.

2.2 *Modally continuous usage*

There is a second feature of the semantics of normative terms that is worth noting at this point. The possible uses of ‘ought’ are what I will call *modally continuous*. This means that for each community that uses ‘ought’ in a particular way, there is a nearby community that uses ‘ought’ differently, but not with drastic differences. That is: there is no possible community where the only nearby communities that use ‘ought’ differently are communities that use ‘ought’ significantly differently. Modal continuity rules out the existence of a world where the only nearby worlds to it are either worlds where normative vocabulary is used in the same way, or vastly differently.

There are several familiar aspects of normative language that support modal continuity. Existing literature has focused on the possibility of normative disagreements between possible communities who use their normative terms differently (see Horgan and Timmons (1992), Horgan and Timmons (2000), and responses). But in addition to these possible disagreements, there would appear to be nothing intrinsic to the use of normative vocabulary that would prevent the existence of small-scale differences of this kind between communities in nearby worlds. If we in fact apply our term ‘ought’ to acts of giving 10% of our income to charity, it could easily have been that we apply our term ‘ought’ only to acts of giving 9.5% instead—perhaps because we are slightly more sensitive to the psychological difficulties of giving away income, and so make some small exceptions in our normative claims for acts that require a certain degree of difficulty. So I will assume, in what follows, that modal continuity is a feature of normative discourse.

Moreover usage of normative terms is not constrained by analytic connections to other terms in our language, or causal connections to properties in the world. While our use of ‘bachelor’ is tied to use of ‘unmarried’ and ‘male’, normative vocabulary does not bear the analytic connections to non-normative terms that would similarly constrain its use (cf. Moore (1903)). Nor are there properties in the world that causally regulate our use of ‘ought’ (cf. Horgan and Timmons (1992)), in contrast with some terms for physical quantities like ‘mass’ and ‘negative charge’. Modal continuity captures a generalization on the absence of these kinds of constraints on possible uses of normative terms.

2.3 *Meta-semantic risk*

Moderate semantic stability and modal continuity in normative terms imply that our normative beliefs are subject to what I will call *meta-semantic risk*. This is the central notion in the epistemological argument against realism that I will investigate in this paper.

At a first pass, meta-semantic risk is present when a term could have referred to a property that is distinct from the property it actually refers to. For example, if ‘ought’ refers to happiness-maximization, one way for meta-semantic risk to be present is for there to be a world where we use ‘ought’ to refer to actions that

maximize happiness and are psychologically feasible for humans to perform. The presence of meta-semantic risk follows from the moderate semantic stability of 'ought'.

By itself, meta-semantic risk is not obviously epistemically troubling. All of our terms, both normative and non-normative alike, would refer to something different given sufficiently drastic changes in our use of the term. (For some non-normative terms, the shift in reference is especially easy to come by: *cf.* our earlier discussion of 'red' and other color-terms.) But the combination of moderate semantic stability and modal continuity that gives rise to the meta-semantic risk in normative terms represents, I will argue, a unique epistemological threat. Since the source of meta-semantic risk in normative terms is a part of the distinctive core of the realist thesis, this represents an epistemological argument that is a fully general problem for realism.

In the next section I will outline the epistemological assumptions that are needed to turn meta-semantic risk into a unique threat to normative knowledge.

3 Epistemic risk

One desideratum for a convincing epistemological argument against realism is that it should avoid relying on a controversial principles about what knowledge or justification *are*. So I will focus here on a simple and commonly accepted necessary condition on knowledge and epistemic justification. This is an *anti-risk* principle: roughly, a belief cannot be knowledge if it is at risk of being false. This principle needs refinement, but the upshot is clear: if realism about the normative is plausible, it then among its consequences must be the epistemic fact that normative beliefs are not at risk of being false. Otherwise realism entails that there is no normative knowledge.

Similar points will apply for epistemic justification. 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 if knowledge is absent because of the presence of a kind of objectionable risk, one will lose justification for a normative belief when one learns that the relevant kind of risk is present.

The relevant notion of *risk* is to be cashed out in terms of what goes on in nearby worlds, or worlds that could easily have obtained.⁸ When one risks dropping one's phone in a pool by standing on the edge of the pool and tossing the phone in the air, this amounts to there being a nearby world where one tosses the phone in the air and it falls into the pool. Likewise when a belief is at risk of being false, this amounts to there being a nearby world in which the belief is false.

I will say that, when a belief could easily be false in the relevant sense, it is subject to *epistemic risk*. Beliefs that are subject to epistemic risk are, by definition, subject to a kind of risk that is incompatible with knowledge. Calling the risk *epistemic* serves to distinguish it from other kinds of risk that may not—or are

⁸Cf. "safety" principles in Sosa (1999), Williamson (2000), and Pritchard (2004).

not by definition—incompatible with knowledge. While I am assuming here that there is a species of risk that precludes knowledge, I am not assuming here that its absence is the *only* condition on knowledge. However it provides a useful picture which illuminates the ways in which mere guesses, beliefs based on the testimony of serial liars, and Gettierized beliefs are not knowledge. And so it provides an especially natural and applicable framework for investigating the relevance of meta-semantic risk to central epistemological notions. A few notes on the types of scenarios in which a belief is subject to epistemic risk are in order.

First, a belief has is at risk only if there are *similar* beliefs that are false in nearby worlds. I can know that I had breakfast 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. (This is not a similar belief, and so does not put my beliefs about breakfast at risk.) But bad companions do not need to be identical to the beliefs they are bad companions for. If one is guessing at the answer to questions about the sums of moderately large numbers, then one's correct guesses won't have nearby worlds where the same belief is false. If one correctly guesses that $634 + 399 = 1033$, then one has a true belief, and moreover this very belief is not false in any nearby world (in all nearby worlds, $634 + 399 = 1033$). So one can't have a false belief that $634 + 399 = 1033$ in any nearby world.

But correctly guessing doesn't 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 (for instance) that that $634 + 399 = 893$. This belief is sufficiently similar to one's actual belief, and since there are nearby worlds where one has false beliefs like this when one is guessing, one's actual true belief that $634 + 399 = 1033$ is at risk.

Here is a second qualification: not all similar false beliefs are bad companions. 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 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 in town). This belief isn't 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.⁹ The reason is that the beliefs are formed by very different processes—specifically, the token causal processes that produce each belief are very dissimilar, as one involves perception and the other involves an abductive inference on the basis of by knowledge of my friend's place of residence. Thus if a belief is at epistemic risk, there must be nearby false beliefs that are both similar in content *and* similar in respect of the token causal processes that produce them.

⁹Cf. Pritchard (2004)

4 Meta-semantic risk and epistemic risk

If our use of 'ought' is subject to meta-semantic risk, then our normative beliefs can be subject to epistemic risk—and hence fail to be knowledge.

The short and ambitious version of this argument is as follows: meta-semantic risk, by definition, is constituted by the possibility of a shift in the reference of 'ought'. In worlds where 'ought' has shifted reference, some agents will have false normative beliefs. So meta-semantic risk implies that normative beliefs are not knowledge.

This short version of the argument is instructive but not convincing on its own. We need to focus on the *ways* in which 'ought' is semantically stable, and the relationship between its semantics and epistemic risk. The needed assumptions are spelled out in the *Argument from Risk*:

- 1 'Ought' is moderately semantically stable and modally continuous.
- 2 If 'ought' is moderately semantically stable and modally continuous, then one could easily be in a world where 'ought' doesn't refer to obligation.
- 3 If one could easily be in a world where 'ought' doesn't refer to obligation, then one could easily have 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.

The first two premises in the Argument from Risk capture the way in which 'ought' is subject to semantic shifts, which are consequences of the core features of realism, as I conceive of them.

Premise 3 connects shifts in the reference of 'ought' to false beliefs about obligation. This is not trivial; it is not generally true that the possibility of a shift in reference entails the possibility of false beliefs about the new subject-matter.

Finally not just any possibility of a false belief implies the absence of knowledge—as we saw in the previous section, only some possible false beliefs represent epistemic risks for our actual belief, which are the kinds that threaten knowledge. Premise 4 claims that the false beliefs which are a consequences of meta-semantic risk are also epistemic risks. This is another substantive assumption.

The connection between meta-semantic risk and epistemic risk the Argument for Risk encapsulates should be troubling for the normative realist. Suppose that our community uses 'ought' to refer to obligation, and moreover that among the obligatory actions are the act of giving 10% of one's annual income to charity. Because of the semantic stability of 'ought', there are many worlds where our use of 'ought' to changes slightly but still refers to obligation—and so refers to a property that has giving 10% of one's annual income to charity in its extension. But the stability is only *moderate*, so there are also worlds where linguistic situation

is one where our use of 'ought' refers to a property distinct from obligation—call it *obligation**—that does not have giving away 10% of one's income in its extension. Moreover 'ought' is modally continuous, which means the world where we refer to *obligation** differs only slightly in how we use 'ought' from a world where we still refer to obligation. We won't know when this shift happens.

One way to fill out the example is as follows. We can suppose, as before, that obligation is the property of happiness-maximization, and this property is a reference magnet. Moreover the nearest reference magnet that is distinct from obligation is the property of being a psychologically feasible happiness-maximizing act. (This is *obligation**.) At a certain point, slight changes in use will produce an overall usage profile for 'ought' that is better fit by *obligation**.

If a typical agent could not feasibly give 10% of her income, then giving 10% of one's income is not in the extension of *obligation**. In a world where 'ought' refers to *obligation**, one speaks falsely if one applies 'ought' to the act of giving 10% of one's income. But just because one is in a community that refers to *obligation**, one might not know this. Because of modal continuity, there is a community which differs ever so slightly from one's own, and refers to obligation. They speak truly when they apply 'ought' to acts of giving 10% of one's income. One can easily be unaware of this shift, and continue to apply 'ought' to the act of giving 10%. But since one in fact refers to *obligation**, one will in typical circumstances come to believe that giving 10% of one's income to charity instantiates *obligation**. This is a false belief—giving 10% of one's income is, by definition, obligatory but not *obligatory**. So being at risk of a semantic shift puts one at risk of having false normative beliefs.

Below I will make the positive case for each premise in the Argument from Risk; in the next section I focus on limitations.

Premise 2: from stability to risk

In the framework I outlined in SS1-2, the aim for a compelling epistemological argument is to show that the core commitments of realism entail that normative beliefs fail to be knowledge, or epistemically justified. The first step in this argument is to show that meta-semantic risk for normative terms entails that a reference-shift could easily have happened. This is what Premise 2 claims.

These reference-shifts imply (if Premises 3 and 4 are true) that we are at risk of having false normative beliefs in the worlds where these shifts occur, and that these false normative beliefs prevent our actual normative beliefs from constituting knowledge. It is worth emphasizing that, while reference-shifts are possible for other non-normative 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; there is no sense in which one color-property is more elite than others and counts as the reference of 'red' even when other color-

properties are better fits. Color-terms are not stable, but rather are semantically *plastic*.¹⁰

But these shifts do not threaten our knowledge of colors—no one could plausibly argue that I don't know that my coffee mug is red on the grounds that my term 'red' could easily have referred to a different property. The 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 believe, of those orangey shades, that they have the property red*, since we are aware that reference shifts with this change in use. So we will not have false beliefs, where semantically plastic terms are involved.

Because of the stability of 'ought', coupled with its modal continuity, we won't be able to have this knowledge of reference-shifts of normative terms in the same way. The stability of 'ought' means that there are a range of worlds where our use changes, but we refer to obligation in each world. But the modal continuity of 'ought' means that, when the reference of 'ought' does shift away from obligation, the use of normative language is only slightly different from worlds where 'ought' continues to refer to obligation. So the reference-shifts for normative terms will, owing to the distinctive features of normative language, be unknowable.

Premise 3: from language to belief

With normative vocabulary, we could easily have been in a world where 'ought' refers to something besides obligation—and not know that this shift has occurred. Premise 3 in the Argument from Risk connects this to the possibility of false *belief*. This is a claim not about language, but about propositional attitudes, including belief, which are a part of our mental lives. So Premise 3 encapsulates the assumption that shifts in the reference of a piece of language have consequences for the content of our beliefs.

In addition the semantic shifts for normative terms will not only have consequences for the content of beliefs; it will also have consequences for the alethic status of these beliefs. Premise 3 claims that if 'ought' shifts reference, some of our normative beliefs will be false. So we need in addition to explain why these beliefs will not only have a different content, but why that content will (in some cases) be false.

Take an example where the term 'ought' shifts reference—say, it refers not to actions which maximize happiness, but rather to psychologically feasible happiness-maximizing acts. This is a fact that is explained by the use of 'ought' by a community of language-users that is better fit by psychologically feasible

¹⁰I am borrowing this terminology from Dorr and Hawthorne (2014).

happiness-maximizing acts than other similarly elite candidate referents. The normative beliefs of language-users in this community will be about psychologically feasible happiness-maximizing actions as well. This is because people will usually rely on the referents of terms in their communal language in order to form beliefs about the world around them. Take for example the central cases that motivate content-externalism for beliefs. Burge's patient with an ailment in her thigh believes she to be arthritis (Burge 1979) has a false, because it is a belief about *arthritis*—which is a condition that affects only the joints. She is able to have this false belief is that she relies on the meaning of 'arthritis' in her public language, and forms a belief about whatever it refers to. In this case the external factors which contribute to the reference of 'arthritis' make possible erroneous views about arthritis.

Things go similarly, I am assuming, with our normative vocabulary. We form normative beliefs the same properties as our peers, and this enables coordination of thought and normative disagreement. We rely on the meaning of 'ought' in our common language. Since normative terms are (as outlined in §2) semantically stable, normative thought will inherit the same stability, and to the same degree.

Thus for example in a world where our usage of 'ought' has shifted just enough to make the term refer to psychologically feasible happiness-maximizing acts, people typically will form beliefs about psychologically feasible happiness-maximization (as opposed to happiness-maximization simpliciter, which I am assuming is identical to obligation). For instance, suppose Suzy forms a belief about the normative status of giving 10% of her income to charity. This is the belief that has the content she would express by uttering the sentence 'one ought to give 10% of one's income to charity.' Since Suzy is in a world where the reference of 'ought' has shifted, this belief doesn't have the content *giving 10% of one's income to charity is obligatory*. Instead she refers to—and forms beliefs about—something different than obligation; her normative beliefs are about psychologically feasible happiness-maximization. Shifts in the usage of normative language by Suzy's community can affect shifts in what Suzy's normative thought is *about*.

Moreover these communal shifts can affect whether Suzy's normative thought's are *true*. In a world where 'ought' has not shifted reference, and refers to the property of happiness-maximization, Suzy's normative belief has the content of her sentence 'one ought to give 10% of one's income to charity'. Moreover she forms a true belief (supposing, as we have been, that giving 10% of one's income maximizes happiness). But now take the world where Suzy's community uses 'ought' slightly differently, and thereby refers to the property of psychologically feasible happiness-maximization instead. 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. Thus she might continue to form beliefs in the same way, believing, as before, the content that would be expressed by the sentence 'one ought to give 10% of one's income to charity'. But in this world, 'ought' refers to psychologically feasible happiness maximization, and giving 10%

of one's income is (we are assuming) not psychologically feasible. Suzy has a false belief, and the only difference between this world where she has the false belief and a nearly indiscernible world where she has a true belief is that some the usage of normative vocabulary by others in her community has changed slightly.

Thus the reference-shifts in Premise 2 can produce false normative normative beliefs in the worlds where the shifts occur.

Premise 4: epistemic risk

The final step in the Argument from Risk is to connect the false beliefs in worlds like these to epistemic risk—the kind of risk of false belief that is incompatible with knowledge. In §3 we gave several features of false beliefs that generate epistemic risk. False beliefs like Suzy's, which are false owing to the phenomenon of meta-semantic risk, bear many of the necessary features of beliefs subject to epistemic risk.

First, we pointed out that possible false beliefs needn't have identical contents in order to put our actual beliefs at risk. The content of these false beliefs only needs to be similar; in such cases similar-but-false beliefs can be enough to prevent our actual true beliefs from being knowledge. This provision is essential for the Argument from Risk to succeed. In worlds where we have false beliefs owing to semantic shifts, the contents of these beliefs are not identical to the content of our actual normative beliefs. Our normative beliefs in the actual world are about obligation, but beliefs in the shifty worlds are about a different property.

Nevertheless these beliefs will be similar enough to put our actual normative beliefs at risk epistemically. If our normative beliefs in the actual world are about happiness-maximization, worlds where we form normative beliefs about psychologically feasible acts that maximize happiness instead will be worlds where our normative beliefs are very similar to those normative beliefs we form in the actual world. So if they are false, they will be beliefs that are candidates to prevent our actual normative beliefs from being knowledge.

Second, only false beliefs that are the results of relevantly similar causal processes will be candidates to put our actual beliefs at epistemic risk. The Argument from Risk focuses on beliefs that plausibly satisfy this condition. Recall the difference between our true normative beliefs and their counterparts which are false owing to meta-semantic risk: the only difference between them need be that the latter are set in a world where the surrounding community of language-users deploys normative vocabulary in a way that is fit best by a highly elite property that is distinct from obligation. This is compatible with nearly everything in the token causal process that produces one's normative beliefs in each world being the same.

Take for example Suzy's false belief about giving 10% of one's income in a shifty world where 'ought' refers to psychologically feasible actions that maximize happiness. The token process that produces this belief—which includes the moral reasoning which underlies the belief, as well as other psychological factors that

can influence moral belief-formation—can be nearly identical in each world. The difference in content—and hence the difference in the truth-conditions for these beliefs—depends on factors that lay outside the causal chain that produces these beliefs. With meta-semantic risk, we need only to change how the surrounding community uses their normative vocabulary in order to describe a world where one’s normative beliefs are false. When Suzy’s normative belief is false owing to meta-semantic risk, the process by which she forms it can be nearly identical to the process by which she forms her actual normative beliefs.

These points together suggest that the case will be quite strong for the conclusion that possible normative beliefs that are false owing to meta-semantic risk will subject our actual normative beliefs to epistemic risk.

5 Limitations of the Argument from Risk

So far I have assembled the credentials for meta-semantic risk as the key notion in an epistemological argument against normative realism. The phenomenon of meta-semantic risk is one that follows from general features of realism, and not some ancillary theses that realists can reject. And it produces possible false beliefs that have the features to make our actual normative beliefs subject to epistemic risk. 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. So the prospects for the Argument from Risk as a general epistemic argument against realism are in some respects quite promising.

But on closer reflection, the conclusion that the Argument from Risk does not deliver this sweeping conclusion. Rather its ambitions will have to be scaled back: while some normative beliefs do fail to be knowledge owing to meta-semantic 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 raise in the conclusion.

5.1 *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 have easily have shifted referents.

Call a world where ‘ought’ has shifted reference a *shifty world*. According to Premise 2, a shifty world is one we could easily have been in, in the following sense: there is *some* world where ‘ought’ refers to obligation but there are only small differences between how we use ‘ought’ in that world and in the shifty world. Moderate semantic stability guarantees the existence of the shifty world; modal continuity guarantees that it is different from the non-shifty world, which we can call a *stable* world.

But Premise 4 claims that the shifty world is close enough to subject our *actual* normative beliefs to epistemic risk. All Premise 2 establishes is that there are *some* stable worlds are such that a small change in usage in them would put us in a shifty world. And these stable worlds will be worlds where Premise 4 of the argument from risk is true: with only a small change in use, we would be in a shifty world, and thereby have some false normative beliefs. So normative beliefs in *these* stable worlds are at risk in a way that prevents them from being knowledge.

But nothing in the premises of the Argument from Risk guarantees that we are in such a stable world. For all we know, ours is a world where 'ought' refers to obligation, and moreover all of the worlds where normative usage is only slightly different from ours are also stable. The argument is at best *contingently* sound—in some worlds Premise 4 will be true, but not all.

Nonetheless this reveals a distinctive epistemic commitment in virtue of the core features of the realist view. Owing to meta-semantic risk, she will need to admit that there is a *risk* that we lack normative knowledge. If we are in a world where a shifty world is nearby—in the sense that, with a small change in use, our term 'ought' would refer to something besides obligation—then we will lack normative knowledge.

5.2 Higher-order normative knowledge

This is related to a kind of skeptical conclusion that the realist must accept for every stable world, including the actual one. If there are some stable worlds that are (epistemically) close to a shifty world, we lack *first-order* normative knowledge. This is a lack of knowledge of truths such as *one ought to give 10% of one's income to charity*. Even if one believes this truth in such a stable world, one won't know it because the belief one forms by an almost identical process in the nearby shifty world is false. We have seen that the Argument from Risk does not guarantee that we lack this first-order normative knowledge.

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. Ours might be a situation where all of the worlds in our immediate vicinity are stable. But this is compatible with some of the worlds in our immediate vicinity, having a shifty world in *their* immediate vicinity. If this is the case, there are worlds in our immediate vicinity where we lack normative knowledge, for familiar reasons.

This doesn't mean that our first-order normative knowledge is threatened, but it does threaten our high-order normative beliefs. We don't know, in some worlds in our immediate vicinity, that we ought to give 10% of our income to charity. But then, in the actual world, when we form the higher-order belief that we *know* that we ought to give 10% of our income to charity, it will be true but at epistemic risk. There are worlds in our immediate vicinity where it is false. 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.¹¹

5.3 *Imprecise normative beliefs*

So far we have seen that the Argument from Risk is limited in what it establishes. 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 meta-semantic 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, in these sense relevant to epistemic risk, there are kinds of beliefs 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 one ought to give 10% 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 small personal embarrassment is impermissible. Even if these beliefs are true, if a 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 cut-off points for obligatory action. For example believing that one ought to give 10% of one's income to charity is incompatible with the belief that one is only obligated to give 9% of one's income to charity. (If the latter belief is true, one is not obligated to give 10%, 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. If 'ought' shifts its reference to refer to acts that maximize happiness and are psychologically feasible for typical humans to perform, then the belief with the content of 'one ought to give 10% of one's income to charity' in the shifty world is false in that world. If most people can't give 10% of their income to charity without significant coercion or psychological distress, then acts of giving 10% don't have the property 'ought' refers to in the shifty world.

But these problems do not necessarily affect non-specific beliefs. Take for instance the belief that one ought to give between 1 and 20% of one's income to charity. (Where this belief is equivalent to the disjunctive belief that one ought either to give exactly 1%, or one ought to give exactly 2%, and so on.) This belief

¹¹We might have to go even further away from the actual world to find worlds where we don't know that we ought to give 10% of our income to charity, owing to the proximity of shifty worlds. Regardless we won't know that we know that we ... that we ought to give 10% of our income to charity for the same reasons; some iteration of knowledge will fail.

has a *margin for error* built in—unlike the specific belief that one ought to give 10% 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 a shifty world.

The belief that one ought to give between 1 and 20% of one’s income to charity does not have a false counterpart in a nearby shifty world. 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 doesn’t believe that one *ought* to give between 1 and 20% of one’s income to charity, since one’s term ‘ought’ doesn’t refer to obligation—one nonetheless forms a true belief anyway. For instance if in the shifty world one’s term ‘ought’ refers to psychologically feasible happiness-maximization, one still forms a true belief, since there is some non-zero percentage of one’s income which is such that donating it is an act that maximizes happiness and is psychologically feasible for typical humans to perform.

Margins for error can protect some non-specific normative beliefs from going false in semantic shifts. This points to a further respect in which the Argument from Risk is limited. 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 Argument from Risk does not threaten all normative knowledge equally.

6 Concluding remarks: the status of epistemological arguments against realism

In §1 I set up the kind of assumptions a successful epistemological argument against realism would need to start from. These include a characterization of the core, defining features of realism, and a simple (relatively) uncontroversial necessary condition on familiar epistemic notions like knowledge and justification.

The Argument from Risk has the features we need for a compelling epistemological argument against realism. But it has limits, some of which were outlined in §5. Where do the limits to the this leave realism’s epistemological credentials?

From one perspective it saves realism from the damning conclusion that realism entails skepticism. The skeptical conclusion is obviously unpalatable, since if it followed from the core commitments of realism that all normative beliefs are not knowledge, we would have strong reason to look elsewhere for a normative metaphysics that makes room for normative knowledge. Even *prima facie* promising epistemological arguments against realism look much less threatening once we focus on versions that threaten to demonstrate that we ought to jettison realism in all its forms.

However from another perspective the Argument from Risk does not leave realism on a par with its competitors when epistemological considerations are in play. The most obvious respect in which this is so follows from the fact that the

argument is at best *contingently* sound, because Premise 3 is true in some but not all worlds. Since higher-order knowledge is absent in all worlds, whether we are in a world where Premise 3 is true or not is something we will be able to tell. So, while skepticism threatens the realist in some worlds, but not others, there is nothing we can do to satisfy ourselves that we are in one of the epistemically friendly worlds. The realist has to admit that, even if normative knowledge is in fact present, we can't know that we have it, or know that we know that we have it, etc. Some may feel that rejecting realism with the goal of finding a more satisfying home for our normative knowledge is a worthwhile project in light of this.

Finally it is worth emphasizing the respects in which settling for knowledge only with respect to normative claims that have sufficiently wide margins for error is not obviously 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% 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 connections. I will close with two examples.

One comes from the purely cognitive side: it is plausible that, when we form beliefs by deducing them from other beliefs, 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-closure*.) Typical agents who hold the belief the wide margin for error that they ought to give between 1 and 20% of their income to charity do so on the basis of a deduction from a specific belief. But if the specific beliefs are not knowledge, then the wide margins-for-error belief that are deduced from them will not be knowledge as well. So by conceding that some normative claims are not known, we risk denying knowledge to other in-principle-knowable claims that are deduced in normal ways.

The other example is from the practical side: it is common to hold that an agent's reasons, in a certain sense, can only be claims that she knows, or has some epistemic access to. For instance if Bob does not know that the police are arriving at the crime scene, it is infelicitous to say that Bob's reason for fleeing the crime scene was that the police were arriving (?). (This is sometimes called an ascription of a *personal* or *motivating* reason.) If Suzy gives 10% of her income to charity out of her sense of moral duty, it is tempting to say that her reason for giving the money is that she is obligated to give 10% of her income to charity. But this can't be her personal reason for acting, if she doesn't know specific normative propositions like this. And Suzy might not be thinking about, much less take the margin-for-error claim to be her reason—most people don't give 10% of their income merely on the basis of the fact that they are obligated to give between 1 and 20% 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. Fully general epistemological arguments against realism may fail to deliver the knock-out blow that some have hoped for, but there is still some room for epistemological considerations to be a part of the overall case against the view.

References

- Boyd, R. N. (1988), How to be a Moral Realist, in G. Sayre-McCord, ed., 'Essays on Moral Realism', Cornell University Press.
- Brink, D. O. (1984), 'Moral Realism and the Skeptical Arguments from Disagreement and Queerness', *Australasian Journal of Philosophy* 62(2), 111–125.
- Burge, T. (1979), Individualism and the Mental, in French, Uehling and Wettstein, eds, 'Midwest Studies in Philosophy, IV', pp. 73–121.
- Cameron, R. (2008), 'Turtles All the Way Down: Regress, Priority and Fundamentality', *The Philosophical Quarterly* 58, 1–14.
- Clarke-Doane, J. (2012), 'Morality and Mathematics: The Evolutionary Challenge', *Ethics* 112(2), 313–340.
- Devitt, M. (1991), *Realism and Truth, Second edition*, Basil Blackwell.
- Dorr, C. and Hawthorne, J. (2014), 'Semantic Plasticity and Speech Reports', *Philosophical Review* 123(3), 281–338.
- Dummett, M. (1963), Realism, in 'Truth and Other Enigmas', Duckworth.
- Dunaway, B. (MS), 'The Metaphysical Conception of Realism'.
- Dunaway, B. and McPherson, T. (2016), 'Reference Magnetism as a Solution to the Normative Twin Earth Problem', *Ergo* .
- Enoch, D. (2011), *Taking Morality Seriously: A Defense of Robust Realism*, Oxford University Press.
- Harman, G. (1986), 'Moral Explanations of Natural Facts—Can Moral Claims Be Tested against Moral Reality?', *Southern Journal of Philosophy* pp. 57–68.
- Hawthorne, J. (2006), Epistemicism and Semantic Plasticity, in 'Metaphysical Essays', Oxford University Press, pp. 185–210.
- Hawthorne, J. (2007), 'Craziness and Metasemantics', *The Philosophical Review* 116(3), 427–441.
- Horgan, T. and Timmons, M. (1992), 'Troubles on Moral Twin Earth: Moral Queerness Revived', *Synthese* 92(2), 221–260.

- Horgan, T. and Timmons, M. (2000), 'Copping Out on Moral Twin Earth', *Synthese* **124**, 139–152.
- Hyman, J. (2015), *Action, Knowledge, and Will*, Oxford University Press.
- Jenkins, C. (2005), 'Realism and Independence', *American Philosophical Quarterly* **42**(3), 199–211.
- Lewis, D. (1983), 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* **61**(4), 343–377.
- Lewis, D. (1984), 'Putnam's Paradox', *Australasian Journal of Philosophy* **62**(3), 221–236.
- Mackie, J. (1977), *Ethics: Inventing Right and Wrong*, Penguin.
- Moore, G. (1903), *Principia Ethica*, Cambridge University Press.
- Pettit, P. (1991), 'Realism and Response-Dependence', *Mind* **100**(4), 587–626.
- Pritchard, D. (2004), *Epistemic Luck*, Oxford University Press.
- Putnam, H. (1981), *Reason, Truth and History*, Cambridge University Press.
- Railton, P. (1986), 'Moral Realism', *The Philosophical Review* **95**(2), 163–207.
- Schroeder, M. (2007), *Slaves of the Passions*, Oxford University Press.
- Shafer-Landau, R. (2003), *Moral Realism: A Defense*, Oxford University Press.
- Sosa, E. (1999), 'How to Defeat Opposition to Moore', *Philosophical Perspectives* **13**, 141–153.
- Street, S. (2006), 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies* **127**(1), 109–166.
- Williams, J. R. G. (2007), 'Eligibility and Inscrutability', *The Philosophical Review* **116**(3), 361–399.
- Williamson, T. (1994), *Vagueness*, Routledge.
- Williamson, T. (2000), *Knowledge and Its Limits*, Oxford University Press.
- Wright, C. (1992), *Truth and Objectivity*, Harvard University Press.