Realism, Meta-semantics, and Risk

Billy Dunaway
University of Missouri–St Louis

Draft of 24th May 2016

0. Preliminaries

Moral realists are often accused of lacking the resources to provide an acceptable epistemology. These allegations take different forms, but examples include the following:

- 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;
- Dworkin's (1996) characterization of the realist as assuming that a quasi-physical quantity called "morons" explain our capacity to form moral judgments.

We could easily add additional examples to the list.[1] But doing so would only make a longer list with a common theme: none of them are very compelling.

In summary the reason is that each argument relies on a characterization of moral realism, or an epistemological principle concerning what would constitute a successful realist epistemology, that the realist can easily reject. Frequently she has plausible independent motivations for demurring: for example contra Harman, mathematical knowledge can provide a model of a domain in which causal interaction is not necessary for knowledge.[2] And contra Street, epistemologists are generally in agreement that the bare *possibility* of forming false beliefs in your situation does not impugn the epistemic status of your *actual* beliefs.[3] Likewise a Dworkinian caricature of the (non-quietist) realist metaphysic as including a commitment to morons is fairly gratuitous.[4]

The aim of this paper is not to repeat these criticisms and replies here. Rather it is to explore the prospects for new ways of developing epistemological problems

[1] For another argument along these lines, see Mackie (1977) for the "argument from queerness".

[2] Clarke-Doane (2012)

[3] Dunaway (Forthcoming)

[4] See Shafer-Landau (2003) and Enoch (2011) for versions of realism that do not traffic in morons.

for realism from secure epistemological and meta-ethical starting points. This requires a detour through some of the finer points of metaphysics and meta-semantics for the realist. It also requires a firm conception of what would constitute success for a realist epistemology (and, in addition, what the conditions of failure look like). Thus some extensive stage-setting is in order before turning to epistemology proper, but the pay-off will be a setting in which epistemological problems for realism can be discussed in a fruitful and general context.

Once the needed clarifications are in order, the contours of the potential epistemological problems I will explore go as follows. Realism is, as a general matter, committed to the possibility of false normative judgment: in a simple case this involves believing the normative proposition *one ought to* ϕ when it is in fact not true that one ought to ϕ. This isn't enough to distinguish realist views from irrealist views, however–see Street (2008). So we need some additional machinery to distinguish the realist from the irrealist.

I outline in Section 1 what I will assume does this additional work. It is a metaphysical thesis according to which the normative is *highly fundamental.* The notion of fundamentality deployed here is the focus of a much literature in recent metaphysics, and can be understood in a number of ways, though for our purposes any number of notions in the area can do the needed work. Moral realism is thus a thesis in moral metaphysics, and I explain in Section 1 how understanding it in this way has a number of theoretical advantages.

Section 2 extends the realist thesis. It is common to supplement the claim that a certain property is highly fundamental with the additional thesis that it is a *reference magnet*. This is a claim in meta-semantics; it holds that the property in question is easy to refer to. What this means is that it is possible to (and, in some cases, fairly easy to) use a term that refers to a property P and yet fail to perfectly track P with one's usage of the term. In the moral case, this means that one might use 'ought' to refer to the normative property *obligation*. But one might fail to track obligation with one's use of 'ought': this amounts to applying 'ought' to some non-obligatory things, and failing to apply 'ought' to some obligatory things. According to the reference magnetism, one's success in still referring to obligation is partially explained by the fundamentality of obligation.

So far these are just theses about the metaphysics and meta-semantics of realism about obligation. The aim of this paper is to assess their consequences for the *epistemology* of realism: that is, for the possibility of having knowledge of the normative, or justified beliefs about the normative. Before turning to examining the consequences of realism for this question, we need to have a concrete conception of what it would take for realism to fail to achieve epistemological success. This should not rest on idiosyncratic epistemological principles that some of the existing epistemological arguments against realism mentioned above rely on. Section 3 identifies such a condition: it is *risk of error,* appropriately qualified. If realism entails that normative beliefs are at risk of error in the appropriate sense, then it will be uncontroversial that realism fails to provide a satisfactory moral epistemology.

Section 4 outlines the argument that the meta-semantics which accompany realism entail the presence of an epistemically problematic risk of error, which I call the Argument from Risk. The argument hinges on a feature of the reference magnetism thesis for 'ought': that reference to obligation will be *moderately stable*. This means that a range of ways of using the term 'ought' will succeed in referring to obligation: small changes in use do not result in reference to a different property. But some changes in use will take us outside the range, and produce reference to a different property. The Argument from Risk claims that the risk of semantic shift entails an epistemically unacceptable risk of false belief.

This is an argument that targets realism on the basis of its distinctive features, so prima facie the anti-realist will not have to confront the possibility that the argument also applies to her view is subject to the same criticism. Section 5 explores the possible implications for realism in light of the argument. Skepticism about moral knowledge is one possibility. But a more nuanced view holds that the realist should think she *might* have knowledge of normative facts, but can never in fact know that she has it. Thus my tentative suggestion is that an epistemological argument against realism can be motivated on secure grounds, but cannot deliver the kind of knockout blow to realism that some anti-realists have hoped for.

1. The Metaphysics of Realism

What is realism about the normative? If the Argument from Risk is going to present a successful epistemic challenge for realism, it must derive epistemic difficulties from the distinctive aspects to realism–and not from features it shares with its anti-realist competitors. (Otherwise, the argument won't present a problem for realism, since its structure would show that the realist's competitors also have some explaining to do.) The paradigmatic example of the realist view is *non-naturalism,* which is articulated in Moore (1903). According to non-naturalism, obligation and other normative properties are sui generis, and are of a metaphysically very different kind than natural properties. But I will assume that some versions of *naturalism* can also be realist views, and so it is worth saying something about what these views have in common.[5]

Naturalist versions of realism include the views in Railton (1986) and Schroeder (2007). According to these views, normative properties reduce to natural properties. For Railton, the natural properties in question concern what one would want, if one were to undergo a process of experiencing various outcomes, updating one's desires based on feedback from the experiences, and consider one's options from an impartial perspective. What one would want, in

[5] Also it is not clear what the content to Moore's own view is, since his primary data-point in favor of non-naturalism is the "Open Question Argument", which shows that one might know what one ought to do in a circumstance, but not know what natural property one's obligation has in that circumstance. But the metaphysical implications of the Open Question phenomenon are dubious because in general one can know a fact without knowing the nature of the fact.

such an informed and impartial state, constitutes what ought to be done. For Schroeder, one's reasons reduce to what would promote one's (actual) desires, and one's obligations are constituted by which of those reasons one has most reason to weight most heavily in deliberation. This is just a rough sketch of each view, but it points to something important about naturalistic versions of realism: it is compatible with reduction, and moreover it is compatible with reduction to broadly mental or psychological properties. Wants and desires–both paradigmatic examples of mental states–feature prominently in each view. But the views are realist nonetheless.

This throws a wrench in one common way to understand the realism/irrealism distinction. According to this approach, realism is the view that the normative is mind-independent, and irrealism is the denial of this. (Street (2006), Devitt (1991)) But these examples suggest that mind-independence is not where the dividing line lies, as some views reduce the normative to mental properties but nonetheless have the trappings of realism.[6] Moreover 'realism' is used with the same sense with reference to other domains, and the mind-independence criterion is clearly inadequate in some cases. For instance a Behaviorist about mental states holds that being in pain is just a matter of exhibiting certain behaviors and having appropriate dispositions to behave. (For example one's being in pain is just the fact that one is screaming, or disposed to ask for Ibuprofin, etc.) Similarly for all other mental states. The Behaviorist view is an irrealist view about mental states. But it reduces mental states to behaviors, which are non-mental entities. So the mind-independence criterion fails to classify the Behaviorist view correctly.[7]

However there are some views of the normative which are both irrealist, and entail that the normative is mind-dependent. While the mind-dependence of the normative can't constitute irrealism, it is in some cases very closely related to it. Take Street's (2008) "Constructivist" view, which is an example of an irrealist view which is also a view on which the normative is mind-dependent. According to this view, to have an obligation to ϕ is just to be such that one would believe that one has an obligation to ϕ if one were fully coherent and in possession of all relevant factual information. Intuitively, this comes out as an irrealist view because obligations are not very objective on this view: I could easily believe that I should ϕ in the relevant conditions, while you continue to believe that you should not ϕ in the same conditions. On Street's view, we are both right: I ought to ϕ, and you ought not to ϕ. So the Constructivist view entails that two individuals in almost identical situations needn't have the same obligations.

By contrast obligation is objective to a much greater degree on developments of versions of naturalistic realism mentioned above. Whereas on Street's view it will be fairly trivial for individuals to have different normative beliefs (after achieving full coherence and factual information) and thereby have different obligations. But on some ways of filling out (with hypotheses that are in part empirical in nature)

[6] For an extreme example of this, see Wedgwood (2007).

[7] See Dunaway (MS) for more detailed discussion.

the views of Railton and Schroeder, the wants that survive feedback loops, or the weightiest desire-promoters, are common across a wide range of agents. In this case obligation will not be highly relative to individuals, as it is on Street's view.[8]

What this implies for the metaphysics of realism is a tricky issue, but here is one plausible hypothesis that fits well with other examples of the realism/irrealism divide outside of ethics. Recall our earlier example of Behaviorism as an irrealist view about mental states. One plausible hypothesis about why this is an irrealist view is the following: mental states like pain are, on this view, highly disjunctive and gerrymandered properties that do not confer much resemblance on objects that instantiate them. After all, there is nothing substantial that screaming and being disposed to ask for Ibuprofen have in common from a Behaviorist perspective. Obligation turns out to be very gerrymandered in a similar way on the Constructivist view. If two agents in almost identical situations (where the differences are only those that are consequences of the agents having different scrutiny-surviving beliefs) can differ in their obligations, then obligation cannot be a very metaphysically robust property.

This is all very picturesque, but it can be sharpened using tools from recent metaphysics. It is common to invoke the language of fundamentality when distinguishing between properties that are very "gerrymandered" and those that are not. Using this language, we can say that obligation is very fundamental on (appropriate developments of) realist views like Railton's and Schroeder's, while obligation is not very fundamental on the irrealist Constructivist view. And more general, realism about a domain is a matter of taking the relevant properties to be very fundamental, while irrealism is a denial that these properties are fundamental to the relevant degree.

There are a number of perspectives on fundamentality and related notions in the metaphysics literature.[9] But the core idea can be captured by pointing to examples of properties that play a certain kind of explanatory role: these include quantities from fundamental physics including mass and charge. These properties would appear in a fundamental, basic account of what the world is like. For instance such an account will not mention that there is someone currently sitting at his computer writing a paper; instead it will describe the mass, charge, and other fundamental physical quantities in the relevant location.

In addition there are some ways of describing the world that do not characterize it at its most fundamental level, but provide a fairly joint-cutting and natural description of how the world is. These are explanatory-but-not-perfectly-fundamental properties like being a molecule and gravitation, and are all fairly fundamental, metaphysically speaking. And they can be contrasted with gerrymandered and gruesome properties like located in Western Russia in the middle third of the 1960s, being composed of my left pinky finger and the tip of the Eiffel

[8] I do not wish to claim that the developments of naturalistic realism which make obligation highly objective are entirely plausible. While remaining neutral on this issue here, I will instead focus on what the distinctive implications of these views if they are true.

[9] See Fine (2001) and Wedgwood (2007)

Tower, and being thought about by me at exactly 3 pm yesterday. One can provide a true description of the world by mentioning these properties, but they are not very fundamental at all. The realist idea, as I will understand it, is that normative properties belong in the former, and not the latter camp.

So the core component of realism about the normative is that normative properties are metaphysically very fundamental. What are the consequences of this for moral epistemology?

2. Magnetism

So far, realism as I have characterized it is just a metaphysical thesis. But there is a very natural thesis that connects the metaphysical status of moral realism with issues concerning normative language and belief. This threatens the epistemic prospects for moral realism.

The thesis in question is known in the literature as *reference magnetism*. Roughly, this is the view that properties which are highly fundamental are easy to refer to. More precisely, begin by taking a community of speakers who use a term t. Their use can schematically be divided into two components: 1. the objects t is applied to, and 2. the other terms t' that are (inferentially or probabilistically) connected to t. For instance: 'red' as used by English speakers is applied (by and large) to red things. Moreover it is inferentially connected to other color-terms: if 'red' applies to something, than 'darker than orange' also applies to it, as does 'lighter than black'. Obviously there are issues of vagueness, mistakes in usage, and the like, but the basic picture is that these aspects of English speakers' use of 'red' contribute to the fact that it refers in English to redness.

One important aspect of this picture is that the reference-determining use of t is community-wide: one individual's usage of t does not by itself determine what t refers to in that individual's mouth. This is evidenced by the fact that members within a community can disagree with each other: vagueness aside, two English speakers might disagree over whether a particular color patch is red. If individual use alone determined what each speaker refers to, then disagreement would be hard to come by: one speaker would refer to a property P that has the color patch in its extension, while the other speaker refers to a distinct property P^* that does not include the patch in its extension. As a result there would be no disagreement: one speaker thinks that patch has property P, while the other thinks it has property P^*.

Reference for t is thus not settled at the individual level; rather it is a community-wide phenomenon. This is because part of the usage of each member of the community includes deference to other members of the community. In other words, each individual in the community has a disposition to coordinate their usage with others, and so the correct referent for a term as used by an individual depends not only on how that individual uses the term, but also on how her peers use it. This isn't to say that one's own meanings are completely outside of one's own control. One's own usage counts for something, and one could always stop deferring. But so long as one is deferential, it will always be possible to make mistakes when applying a term.

But a whole community's usage is still not enough to determine a referent for t. There can be disagreements between communities who do not defer to each other at all. One community might apply t to a property P, while the other community applies t to a distinct property P^* has a different extension. If community-wide use settles reference, we have no disagreement over cases that are in the extension of P but not P^*: one community thinks that the case has property P, while the other thinks it lacks property P^*. There is no disagreement between the two communities.

So the communal aspect to pure use of a term is not enough to determine a referent. In addition the machinery of reference-determination includes some factors that extend outside the patterns of linguistic usage by a community.

Here it will be useful to contrast cases where extra-communal reference-determining elements are needed, from cases where they are not. Here are a few:

Gold: Communities C and C^* use the term 'gold' to refer to metallic substances that they treat as valuable. C applies 'gold' to all and only samples of the element Au, but C^* applies 'gold' not only to samples of Au but also to some instances of iron pyrite ("fools gold"), which they treat as indiscernible from Au.[10]

Red: Communities C and C^* use the term 'red' to refer to colors, which are surface reflectance properties. C applies 'red' to surfaces that reflect light wavelengths between 620?750 nm. (This is the conventional demarcation of the red spectrum.[11]) C^* applies 'red' to surfaces that reflect light wavelengths between 620?750 nm, but also to surfaces that reflect wavelengths from 600?620 nm. (Thus C^* applies 'red' to some colors in the red-orange range as well.)

Charge: Communities C and C^* both use 'charge' to refer to substances with physically measurable properties. C applies 'charge' to items that magnetically repel each other, while C^* applies 'charge' to mutually repellant substances, but some substances that don't repel each other either.

Salad: Communities C and C^* both use 'salad' to refer to edible dishes traditionally served at dinner. C applies 'salad' to dishes composed primarily of lettuce only. C^* applies 'salad' to lettuce-based dishes, but also to dishes primarily made of spinach, corn, cabbage, or couscous.[12]

[10] See Dunaway and McPherson (MS)

[11] `https://en.wikipedia.org/wiki/Visible_spectrum`

[12] See Dorr and Hawthorne (2014)

There is an intuitive distinction between speakers from C and C^* in Gold and Charge, on the one hand, and Red and Salad on the other. Imagine that representative speakers from C and C^* meet each other, and without deference discuss a case over which their respective usage differs. If speakers from Gold are discussing a piece of iron pyrite, and the speaker form C* points and says "that is gold" while the speaker from C points and says "that is not gold", they intuitively disagree, which is to say both of their utterances cannot simultaneously be true. Similarly for speakers in Charge: if a representative speaker from C^* points at a neutron and says "that is charged" while a representative speaker from C points at the same neutron and says "that is not charged", they disagree with each other in the same way.

But Red and Salad show that this kind of disagreement will not always be present. Representative speakers from Red will not disagree when the speaker from C^* points to a red-orange patch reflecting light wavelengths of 610 nm and says "that is red" while the speaker from C points at the same patch and says "that is not red". Likewise representative speakers from Salad will not disagree when the speaker from C^* points at a couscous-based dish and says "that is a salad" while the speaker from C points at the same dish and says "that is not a salad".

The immediate lesson about reference-determination to take from these cases is that sometimes, but not always, features external to a community's usage will conspire to fix the referent of a term in a way that diverges from that community's pattern of application of that term. In Gold and Charge, the divergent use of members of C^* doesn't change the referent–these speakers still refer to gold and charge, and so disagree with speakers from C since their statements express incompatible propositions in the same context. These terms are *semantically stable,* since the difference in use from C and C^* does not generate a difference in referent.[13]

But in Red and Salad no such disagreement takes place. When speakers from C^* use 'red' and 'salad' differently, they do succeed in shifting the referent of these terms. Hence they refer to a property that includes surfaces that reflect 610 nm wavelengths, and a property that applies to dishes that are couscous-based. Since propositions that ascribe this property to the relevant items are compatible with the proposition expressed by members of C in the same context, there is no disagreement. These terms are semantically plastic, since the difference in use between C and C^* produces a different referent.

The natural explanation for the difference in semantic stability and plasticity between these cases is reference magnetism.[14] Since the element Au and negative charge are highly fundamental properties, they can serve as magnets: even when a community does not apply their terms 'gold' and 'charge' to these properties without fail, such a community will succeed in referring to gold and charge if the divergence in use is not too large. These properties are easy to refer to. But

[13] Dorr and Hawthorne (2014), Schoenfield (forthcoming)

[14] See Lewis (1983, 1984) for more work for a notion in the vicinity.

the properties of reflecting light wavelengths between 620 and 700 nm, or being primarily composed of lettuce (rather than spinach or couscous), are not very fundamental by comparison. So they are not reference magnets, and it is not very easy to refer to these specific properties rather than distinct properties in the vicinity.

There are two arguments that moral realists should think that normative properties are highly fundamental reference magnets.

The first is the *Argument from Realism*: this argument proceeds from the general character of realism in the previous section, applied to the normative domain. If realism about an arbitrary domain is the thesis that the relevant domain is highly fundamental, then realism about the normative is the thesis that the normative is highly fundamental. Then, since reference magnetism applies to highly fundamental domains in general, it applies in particular to the normative domain. Normative properties are reference magnets.

The second is the *Argument from Disagreement:* just as communities who use their term 'gold' in different ways can nonetheless succeed in referring to gold, likewise communities who use the normative term 'ought' will nonetheless succeed in referring to the same thing. Obligation is semantically stable, and so it is a reference magnet. The evidence for this is the pervasiveness of disagreement: take two communities who use 'ought' in different ways, one by applying it to all and only actions that maximize happiness, the other by refraining from applying it to happiness-maximizing actions that infringe on a rational agent's autonomy. These communities will disagree.[15] The analogous explanation is that 'ought' is semantically stable because obligation is a highly fundamental reference magnet.[16]

While normative terms are semantically stable, they are not *radically* stable. It is not impossible to imagine communities who use their normative terms in systematically different ways, and thereby acquire different referents for their terms. Here is one (extremely simple) example: we can imagine a community who uses the normative 'ought' exclusively for actions that are best, all-things-considered, to perform. Another community might use 'ought' exclusively for actions that satisfy a kind of *bounded optimality* constraint: they are the best, among the actions that can feasibly be performed.[17] In this case it is natural to interpret each community's use of normative language as accurate, and referring to different properties. (It is also natural to treat the community as not disagreeing with each other.) Of course each term will still be fairly stable: slight departures from perfect usage in either community will not produce a new referent. But the fact that wholesale shift in usage from one community to the other can produce a change in referent shows that 'ought' is only *moderately* semantically stable: reference magnetism protects against frequent semantic shifts in the reference of

[15] Cf. the 'Moral Twin Earth' thought experiment from Horgan and Timmons (1991, 1992*a*,*b*)

[16] This is developed more in Dunaway and McPherson (MS)

[17] Of course ordinary English has a context-sensitive 'ought' that can express either of these notions in different contexts (Cf. Kratzer (1977)). I will ignore this complication here for the same of exposition, though the points made here could be accommodated within a contextualist framework.

‘ought’, but it does not guarantee that they are absent entirely.

I won’t defend the arguments for the moderate semantic stability of ‘ought’ in detail here. Instead here I will explore the epistemological consequences of the position, if it is true. This will provide a framework for an epistemological argument against realism that does not rely on optional features of the realist view, but rather proceeds from the core of the view itself. In the next section I spell out the epistemological principles that, while fairly uncontroversial, seem to spell epistemic trouble for realism’s commitment to the moderate semantic stability of ‘ought’.

3. Risk

A convincing epistemological argument should avoid the controversial principles employed by many existing epistemological arguments against realism. So instead of relying on epistemological constraints that cannot be independently motivated, I will focus here on a simple and commonly accepted necessary condition (or is at least proxy for a conjunction of necessary conditions) 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.

But the semantic stability of normative terms threatens to put normative beliefs at risk of being false. So the core tenets normative realism appears to put normative beliefs at risk of epistemically disastrous error.

First, a closer look at what kind of risk of falsity is at issue. *Risk* is to be cashed out in terms of nearby worlds, or worlds that could easily have obtained.[18] When a belief is at risk of being false, this amounts to there being a world that could easily have obtained (a “nearby world”) in which the belief is false. 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.

Call the nearby false belief that puts one’s actual belief at risk and destroys knowledge a *bad counterpart*.[19] Beliefs that are not knowledge have bad counterparts. This picture illuminates the ways in which mere guesses, beliefs based on the testimony of serial liars, and Gettierized beliefs are not knowledge.[20] But it needs refinement.

First, a belief has a bad counterpart (in the relevant sense) if there are *similar*

[18] Cf. Williamson (2000): here it is important to emphasize that the relevant sense in which a world “could easily” have obtained may ultimately be understood in partially epistemic terms. The anti-risk principle should not be understood as ipso fact offering a reductive analysis of (a component of) knowledge.

[19] Cf. Dunaway and Hawthorne (Forthcoming)

[20] This is a common idea in “safety” constraints on knowledge; see Sosa (1999), Williamson (2000), and Pritchard (2004).

counterpart beliefs that are false in nearby worlds. Bad counterparts do not need to be *identical* to the beliefs they are bad counterparts for. If you are guessing whether a fair coin will land heads when flipped, and successfully guess that it will land heads, you nevertheless don't know that it will land heads. Your belief that it will land heads in the nearby world where it lands tails is a bad counterpart. But take an analogous case: one is guessing at the answer to questions about the sums of moderately large numbers. If one correctly guesses that 634 + 399 = 1033, then one has a true belief. But 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 necessary truths 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 that that 634 + 399 = 893. This belief is sufficiently similar to one's actual (true) belief that it can serve as a bad counterpart, and thereby prevents it from being knowledge.

Here is a second amendment: not all similar false beliefs are bad counterparts. Suppose I see you walking past my office door, and come on that basis to believe that you are in town. I can know that you are in town. But I could easily have believed that you are not in town: suppose in addition that I know that you live in a different city, and I have no other evidence that you are in town. I could easily have not looked out my office door at the moment you walked by (or I could easily have left the door closed), and believed on the basis of my knowledge our your typical place of residence that you are not in town. This is a nearby false belief. But it is not a bad companion in the present sense: it doesn't destroy the status of your actual true belief as knowledge.[21] The reason is that nearby beliefs aren't candidates for bad companionship if they are formed by very different token processes. Here my actual true belief is formed on the basis of my seeing you walk by my door, whereas the would-be bad companion is formed on the basis of general knowledge about your usual residence. The difference in the causal process that produces the nearby belief prevents it from threatening the status of my very near belief.

So beliefs that have bad companions are at risk of being false in an epistemically relevant sense. They are not knowledge. This is a fairly uncontroversial anti-risk principle in epistemology, supplemented with a few tweaks to accommodate the finer details of the kind of risk that is epistemically relevant. Do the commitments of realism about the normative, as outlined in the previous section, entail that normative beliefs have bad companions?

There is reason to think that they do. This can be spelled out in the following schematic argument, which I will call the *Argument from Risk:*

Argument from Risk

1 'Ought' is moderately semantically stable.

[21] Cf. Pritchard (2004)

2 If 'ought' is moderately semantically stable, 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 have bad companions, and are not knowledge.

The premises in the Argument from Risk jointly imply that normative beliefs are not knowledge. Premise 1 is a re-statement of the metaphysical and metasemantic commitments of normative realism, which we outlined in the previous section. Premise 4 summarizes the requirement that knowledge requires the absence of risk of false belief. Premises 2 and 3 connect the metasemantic and metaphysical features of realism with its alleged epistemic failures. The rest of this paper will be primarily concerned with whether they are true, and whether the true readings of these premises all rely the same reading of the expression 'could easily have been'.

One final point is in order before we dive into the substantive implications of the Argument from Risk.

It might be natural to think that a simpler route is available from moderate semantic stability to epistemic failure. If normative terms are moderately stable, then it is possible for one to apply 'ought' to an action that is not obligatory. Since 'ought' is semantically stable, the mere fact of one's use does not guarantee that 'ought' refers to a property that is instantiated by all of the actions one applies 'ought' to. This is sufficient for the possibility of false normative beliefs–and it appears uncontroversial that such error is possible on the realist view.

There are several reasons why this direct argument for the possibility of error is not threatening to the realist.[22] Of course some ways of forming normative beliefs will produce genuine bad companions–beliefs in similar but false claims, formed in nearby worlds, by a sufficiently similar token causal process. For instance if one does the analogue of guessing mathematical claims, but with normative propositions instead, the procedure will produce bad companions, even if one happens to guess the normative truth. Supples as an illustration, that one correctly guesses that a particular murder is wrong. If one is guessing, then one could easily have been in a world where one falsely believed that the murder in question is not wrong. This belief is highly similar to one's actual belief, and it is formed by a very similar token process (guessing). So it is a bad companion; one doesn't know that the murder in question is wrong.

[22] Even paradigmatic anti-realist views acknowledge, and even embrace, the possibility of making false normative judgments (see Street (2008)). So even if the possibility that one makes different (and false) normative judgments entailed normative skepticism for the realist, this wouldn't necessarily constitute a disadvantage for the realist, if competitor views were committed to the same result. I will not pursue this parity argument further here, however, since there are compelling independent reasons to think that the possibility of this kind of error does not threaten skepticism for the realist.

But this kind of bad companionship doesn't arise for more typical methods of normative belief-formation. Take someone who witnesses a murder (or imaginatively appreciates the normatively relevant features of the murder) and comes to believe that it is wrong. It is true that, had things gone differently, the same individual, considering the same murder, could have come to believe that it is not wrong. But this conclusion couldn't have come about by a token process that is very similar. One could come to the erroneous conclusion if one wasn't paying careful attention when presented (via perception or imagination) with the relevant facts about the murder. Or, one could have been raised with a very different moral sensibility toward the type of murdering in question. These are familiar possibilities, and I won't argue that they aren't also nearby possibilities here. But that won't suffice for bad companionship, since the token processes producing the false beliefs won't much resemble the process which produced one's actual (true) belief that the murder in question was wrong. The realist needn't worry about the bare possibility of error in this case.

We need to be restrained when drawing from this example general lessons for the realist's capacity to accommodate all cases of nearby changes in normative belief. It isn't obvious that, in every case where there is a possibility of nearby error, the same avenue to blocking the conclusion that a bad companion exists will be available. Whether (and to what extent) the strategy deployed above can help the realist save normative knowledge is an open question.

Fortunately we needn't settle this question here. The fact that there is a strategy for the realist that saves *some* cases of normative knowledge shows there isn't a structural incompatibility between realism and the existence of normative knowledge. How far that knowledge extends is an open, and interesting question. But this is not the place to look for an epistemological argument that strikes a blow to the overall prospects for realism. For this kind of argument, we need to look at alternative ways in which bad companions can enter into our epistemic lives.

4. Evaluating the Argument from Risk

Realism about the normative implies that normative properties are reference magnets and are therefore easy to refer to–in particular, they are moderately semantically stable. The Argument from Risk claims that this has drastic epistemological consequences: normative beliefs have bad companions, and therefore lack one of the necessary conditions on knowledge.

We learned at the conclusion of the previous section that the possibility of one's normative beliefs changing to false beliefs owing to the stability of 'ought' does not constitute a general epistemic threat to realism. More worrying is the opposite problem: the moderate stability of 'ought' guarantees that one's beliefs could be false, not because the beliefs themselves change, but because their content changes. This is the avenue that the Argument from Risk pursues, and it will be fruitful to look in detail at the premises 2 and 3 to draw some substantial conclusions about the epistemic prospects for realism.

Premise 3: reference shift to falsity

I will take these premises in reverse order. Premise 3 says:

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.

The central motivating idea behind this premise is a thesis about the connection between language and thought. One will typically token beliefs about the properties which one has lexical items to refer to. It is perhaps possible in principle that could believe, of a property O, that it is instantiated by an action a, even though one doesn't have a term in one's language that refers to O. But for our purposes what matters here is that if in one language one has a term that refers to O^*, and one applies that term to a, then one will usually in addition come to believe that a has O^*.

When one's normative language undergoes semantic shifts, the connection between language and thought can give rise to false normative beliefs. Suppose one applies one's word 'ought' to a specific action that is obligatory–for example, one applies 'ought' to the act of Alice's giving 10% of her income to charity, in her concrete situation (which includes everything from income level, financial responsibilities, and probable effects of her donations). If one's word 'ought' refers to the property of obligation (which, by hypothesis this specific action available to Alice has) then the connection between language and thought will give rise to a true normative belief.

But suppose in addition that one's linguistic community's usage has shifted. There are many ways for this kind of shift to happen, where the shifted referent is a property the action of Alice's giving away 10% of her income does not have. Here is one: suppose Alice's community starts to use 'ought', with sufficient uniformity and consistency, only for actions that are psychologically feasible for typical agents to perform. The community in effect requires "obligatory" actions to be those normal agents could perform without significant emotional distress, and in the absence of external coercion or bribery. And let's suppose a typical agent like Alice could not feasibly give 10% of her income in her situation. (If that sounds implausible, fill in the details of Alice's situation appropriately, or find another area where morality appears to be very demanding.)

'Ought' in the mouths of this community refer to the property that has the intersection of obligatory and psychologically feasible acts in its extension. Call this property *obligation**;* I will suppose for the sake of illustration that obligation* is distinct from obligation.

If one is part of a community whose usage of 'ought' as a whole has semantically shifted to refer to obligation*, it needn't be that one's *own* usage has shifted. One might continue to apply ‘ought’ to all and only actions that instantiate obligation, but nonetheless have one’s referent shift to obligation*, in virtue of changes in one’s linguistic community. This is an artifact of the community-wide aspect of reference-determination. By hypothesis, some actions that are obligatory do not instantiate obligation*. But one will apply ‘ought’ to them, and by the connection between language and thought, will frequently form the corresponding beliefs. So one will have false normative beliefs. This is, in schematic form, the argument for Premise 3.

It is worth highlighting the those aspects of reference-determination and belief-formation generate that false beliefs in this scenario.

First, the fact that reference is determined by the usage of ‘ought’ by a linguistic community *as a whole* is crucial here. We can get semantic shifts for ‘ought’ simply by virtue of our linguistic peers’ changes in usage–the actions we apply ‘ought’ to might remain perfectly constant pre- and post-shift. Deference to others can have the unfortunate side-effect of the semantic rug getting pulled out from under us.

Second, the moderate stability of ‘ought’ induced by reference magnetism causes *significant* shifts in reference. Jumping from reference magnet to a different reference magnet will impugn many beliefs: a wide range of important beliefs that come out as true on the first reference assignment will be false on the second. By way of contrast, if ‘ought’ were not somewhat stable owing to magnetism, changes in community-wide use would generate frequent but quite small changes in reference; in this case the term would be *semantically plastic*.[23] While plasticity would involve changes in meaning, the consequences for the veracity of the corresponding beliefs is much less threatening: small changes in reference do not threaten the truth of all but one’s most specific beliefs. While the right way forward to diffuse the epistemic consequences of plasticity is not a straightforward matter in itself, I will not dwell on it here since letting the problem of plasticity go unresolved would lead to widespread skepticism.

Third, one’s beliefs about matters normative needn’t undergo any explicit change to go false. On the belief-formation side, one might explicitly and sincerely accept the sentence ’I ought to ϕ’ in both a world where ‘ought’ refers to obligation, and a world where the word semantically shifts to refer to a different property, which is directed toward ϕ-ing. On the basis of explicitly and sincerely accepting this sentence, one will thereby come to a belief by almost identical processes in both worlds. Of course the beliefs one arrives at by this process will be different–one is about obligation; the other is about a distinct property.

This raises a final fourth and final point about normative beliefs when ‘ought’ semantically shifts. This is that there is a substantive sense in which the beliefs formed by tokening the sentence ’I ought to ϕ’ in worlds pre- and post-semantic shift are both *normative* beliefs. They are not about the same property (Cf. the third classificatory point above) but they are still substantially similar since they play

[23] Cf. Dorr and Hawthorne (2014) for terminology and more discussion.

the functional role of normative beliefs in both cases. That is: they are connected to motivation, the attitudes of blame and praise, and obey the same requirements of consistency and universality. Thus I will say that the beliefs are normative even when 'ought' has undergone semantic shift.[24]

These details bridge some of the gap between a nearby false beliefs, and the more troublesome existence of bad companions for normative beliefs. Since the beliefs pre- and post-semantic shift are all normative beliefs, the false beliefs induced by the semantic shift will be sufficiently similar. Moreover the process by which the beliefs are formed are nearly identical–the only change that produces the false belief is a change in the usage of the surrounding community. So if false beliefs of this kind will satisfy all of the criteria for bad companionship if they occur in nearby worlds.

Premise 2: Stability to reference shift

So one can acquire a false belief in virtue of the reference shift of a moderately stable term like 'ought'. Is this belief a bad companion for one's actual normative belief? If the reference shift happens in a nearby world, it is.

Premise 2 claims that the nearbyness of such false beliefs follows from the moderate stability of 'ought':

2 If 'ought' is moderately semantically stable, then one could easily be in a world where 'ought' doesn't refer to obligation.

We should note at the outset that there is a sense in which Premise 2 is clearly true. Worlds where 'ought' has shifted in reference, and does not refer to obligation, can be *phenomenally indistinguishable* from worlds where 'ought' still refers to obligation. This is a consequence of the communal contribution to reference-determination: since reference depends on usage of one's community, and changes in communal usage needn't present themselves in the phenomenology of a competent speaker, the occurrence of a semantic shift needn't make a difference to what a competent speaker is phenomenally aware of.[25]

So in many situations one could easily have a false belief owing to a semantic shift in 'ought', in the sense that one can't phenomenally discriminate between states where a shift has occurred and states where it hasn't. If someone asks you to cite an aspect of the phenomenally presented world as evidence that a shift hasn't occurred, in most situations one will be forced to respond, "I can't point to any conclusive evidence; a shift very well might have occurred."

[24] This label isn't important. If one wishes to reserve the term 'normative' for terms that refer to obligation and related properties, then the subsequent discussion can be rephrased without the term. What matters is that the beliefs are substantively similar to each other in virtue of the functional role they play–if this role isn't labeled "normative", nothing of consequence follows from this fact.

[25] In addition, competent speakers might not know when a change in usage generates a semantic shift, since they needn't know the specific character of the supervenience relation between meaning and usage facts. See Williamson (1994) on this point in relation to vague terms.

Of course this kind of nerabyness–call it "phenomenal closeness" since worlds that are close in this sense are worlds that cannot be phenomenally distinguished from the actual world–is not the only sense we might give to the kind of closeness that matters for bad companionship. There are clear cases where a phenomenally close world is not a world one could easily have been in in any salient sense. For instance take a highly skilled tennis player with a large lead at the end of a match against an inferior opponent. In this setting it is highly natural for the player to say to herself, "there is no risk that I won't win the match"–that is, that there is no nearby world where she doesn't with. The fact that there are worlds where she is a brain in a vat and has the same phenomenal experiences yet does not win, does nothing to cast doubt on the fact that she is not at risk of not winning. Phenomenal closeness of a world does not entail closeness in the sense relevant to the tennis player's concerns.[26]

Phenomenal closeness is likewise not central to the concerns of epistemology. Someone who is in a normal perceptual environment and is at no risk of being envatted is not at risk of having false perceptual beliefs. But they cannot phenomenally distinguish their actual environment from one in which they are envatted. Likewise someone who has been told by a knowledgeable interlocutor that there is a traffic delay on the highway is not putting herself at risk by believing that there is a traffic delay, even though nothing in her phenomenology distinguishes the interlocutor from someone who is a habitual liar. These cases can be easily multiplied: the sense in which bad companionship involves the nearbyness of a false belief does not require that all worlds where one has a false belief can be ruled out on the basis of phenomenally presented evidence alone. Such an approach would be much too strict, ruling out knowledge of all but the most trivial truths.

So the moderate semantic stability of 'ought' plausibly suggests that worlds where a semantic shift occurs are phenomenally close. In these shifty worlds, one has false normative beliefs. So worlds where one has some false normative beliefs are phenomenally close.

But this doesn't guarantee that these worlds are close in the sense required to make Premise 3 guarantee the presence of bad companions for one's actual normative beliefs. Phenomenal closeness does not guarantee bad companionship. Premise 2 might still be true when we fix on the relevant reading of closeness, but focusing on the phenomenal closeness will not be helpful for establishing this fact.

Suppose then that we fix on a reading of "close" in Premise 2 on which close worlds are, by definition, those that are close in the sense required for bad companionship. Call this *epistemic* closeness. This will ward of an understanding of the Argument from Risk on which the premises are all true, but rely on equivocal readings of "could easily be" (and cognate expressions) in order to

[26] Or, the fact that a schemer might have just manipulated the balls in favor of her opponent by implanting tiny mechanisms that allow the balls to be remotely manipulated while in play (and since the manipulator favors the less skilled player, the schemer has thereby greatly increased her chances of winning).

secure their truth.

Understood in terms of epistemic closeness, Premise 2 is not necessarily true. That is: at every world of evaluation, the material conditional 'if 'ought' is moderately semantically stable, then one could easily be (in the sense tied to epistemic closeness) in a world where 'ought' doesn't refer to obligation' is not true at every possible world. These worlds are those in the "center" of the sphere of worlds where, owing to the moderate semantic stability of 'ought', the reference of 'ought' is the same throughout every world of the sphere. In these worlds, one is as it were not at risk of leaving the sphere.

More precisely: the moderate semantic stability of 'ought' guarantees that there is a set of worlds, M, where at every world in M, 'ought' (as used by one's community) refers to obligation. Moreover the worlds in M will be fairly continuous in how one's community uses 'ought' in those worlds: the worlds form a "sphere" in the sense that for every world in M, there are other worlds also in M where the community usage of 'ought' differs only in minimal respects. (In other words: there is no world in M where, in order to get to another world in M, one has to jump to a world where the community usage of 'ought' is substantially different in order to do so.) Of course there might be pairs of worlds in M where the community usage facts in the worlds are substantially different–what the sphericality of M amounts to is the claim that such worlds will be connected by a chain of worlds, also in M, where usage facts in each world in the chain are minimally different from the usage in its neighbors.

Some worlds in the M-sphere are at the center: one could not easily have been in a world that is not in M; which is to say every world one could easily have been in will also be a world in M. (If we reserve the label E_w for the set of worlds that are epistemically close to w, this amounts to the claim that some world w^* in M is such that every world in E_{w^*} is also in M.) These will be worlds at which Premise 2 is false. So it is not necessarily true. Once we fix on a reading of Premise 2 on which closeness is read as epistemic closeness, the Argument from Risk contains a premise that is possibly false.

But the premise is not necessarily false. Premise 2 is evaluated at a world which is at the edge of the sphere M, there are guaranteed to be some worlds which are epistemically close and which 'ought' has not shifted reference. But there will be other worlds at which it has shifted–if we find ourselves in a world which is in a precarious position at the edge of M, we can find another world within the bounds of epistemic closeness for the precarious world, but which do not fall within the boundaries of M. (If w' is a world at the fringes of M, there will be some worlds in $E_{w'}$ which are not in M.) The epistemic evils threatened by the Argument from Risk will be very real in these worlds.

It is important to be cautious about how far these evils spread. The mere fact that there are some worlds in M which are epistemically close to worlds outside M does not entail that all worlds in M are epistemically close to worlds outside M. This is a purely formal point: being close to being close to a world w does not entail being close to w. The point can be made in terms of risk: I might go to a party where there is a risk, but no guarantee, that I will gamble \$10. And if I gamble the \$10, then there is the risk that I will lose it. But just by going to the party I do not thereby risk \$10: I might have gone and not gambled at all, and hence at no point put my money at risk.

There are some cases which appear to show that sometimes risk of a risk collapses into risk simplicter. For instance Dorr and Hawthorne make the following offhand comment:

> If aliens are in fact going to invade the Earth in 2050, it would be rather tendentious to claim that we can still know that they will not invade in the next year. (Dorr and Hawthorne 2014: 321)

Here we can think of worlds where aliens invade in 2049 as epistemically close to worlds where they invade in 2050; worlds where they invade in 2048 are epistemically close to worlds where they invade in 2049, etc. Assuming they in fact invade in 2050, then we can't know in 2049 that they won't invade, since there is an epistemically close world where that belief is false. But it also seems that we can't know that they won't invade next year, and so the natural conclusion to draw is that being in a world that is close to a world that is close to where the invasion happens in 2050 is sufficient for being close to the invasion.[27]

What this case suggests is not that some iteration of closeness is sufficient for epistemic closeness. Rather it shows that we need to careful about how we circumscribe the worlds that are epistemically close simplicter. Just because in fact the aliens will invade in 2050, it doesn't follow that a world where they invade next year is distant. (Separation by 33 years does not necessarily generate separation by 33 closeness relations.) The case as described doesn't suggest any particular mechanism by which the actual invasion date is decades after one forms the relevant belief. So for all the description of the case says, the invasion could easily happen next year. Since the world where this happens is a world where one has a false belief, one doesn't know that the invasion won't happen next year. Getting to this conclusion does't require collapsing iterations of closeness.

So the Argument from Risk is possibly but not necessarily sound, owing to the contingent truth of one of its premises. On the reading of Premise 2 which is required to ensure the validity of the Argument from Risk is false in some worlds and true in others. The failure of iterations of closeness to collapse into closeness simpliciter guarantees that, owing to the moderate semantic stability of 'ought', not every world where 'ought' refers to obligation will also be a world where it could easily have referred to a different property. In closing I will explore the consequences of this for the epistemology of moral realism.

[27] This principle would reduce any number of iterations of closeness to a single iteration by repeated applications.

5. Conclusion

The first conclusion to focus on is to some extent a worrisome epistemic result from realism: if we are in a world close to a semantic shift for 'ought', our normative beliefs will fail to be knowledge. The Argument from Risk will be sound in these worlds. So depending on where the world we are in is located in modal space, we might fail to have normative knowledge. And regardless of where we are actually located, the realist is forced to acknowledge that she faces an additional obstacle to explaining normative knowledge that her anti-realist counterparts need not face.

Moreover whether this obstacle can be overcome is entirely a matter of the external environment we find ourselves in. It isn't as if the realist faces this hurdle because she construes normative facts in a way that requires agents to make extra cognitive effort in order to get in touch with them. Rather the relevant epistemic difference between realism and anti-realism is just that the realist faces the possibility of being in a world where she makes no mistakes in her process of forming normative beliefs and making inferences with them, but still fails to form normative knowledge simply because a semantic shift looms nearby.

But this is not a wholesale rejection of the possibility of normative knowledge for the realist. Skepticism is not a guaranteed consequence of the view. If we find ourselves at a sufficient distance from a semantic shift, knowledge of normative propositions is not under threat.

A second and more fundamental question for the realist is whether we can *know* we are in one of the knowledge-conducive worlds. Here there is a strong case that we cannot, though I will not argue that the case is decisive. Call the worlds where 'ought' both refers to obligation, and in addition are not epistemically close to worlds where the referent of 'ought' has shifted, *safe worlds*. In safe worlds, one can have normative knowledge.[28] But suppose in addition in one of these worlds that one believes that one knows some relevant normative propositions. For instance: consider a safe world where it is true that I ought to ϕ, I know that I ought to ϕ, and I believe that I know that I ought to ϕ. Even though the belief that I ought to ϕ has no bad companions (recall that we are in a safe world), the belief that I know I ought to ϕ does.

Since we are in a safe world, we are not epistemically close to worlds where we have false normative beliefs owing to a semantic shift. There are other non-safe worlds where the shift has not happened, but which are epistemically close to worlds where the shift has occurred. Call these *unsafe worlds*. We are not in an unsafe world, but we are epistemically close to some. But in unsafe worlds it is false that we have normative knowledge. That is, even if it is true in such a world that I ought to ϕ, is is false that I know that I ought to ϕ. So if I believe that I

[28] How can we be sure that there are safe worlds? This requires the range of worlds in which the reference of 'ought' remains fixed to extend beyond the epistemically close worlds. Nothing I have said here provides definitive evidence for this assumption. If we could show that it is unjustified, the claim that the Argument from Risk is sound in every world would be up and running again.

know that I ought to ϕ in an unsafe world, I have a false belief. Moreover such belief will satisfy all of the conditions for bad companionship with the belief that I know that I ought to ϕ in a safe world: it will be sufficiently similar, formed by a nearly identical process, and in an epistemically close world. So I can't know that I know I ought to ϕ, even in a safe world where first-order normative knowledge is available.

There are many delicate issues here. But instead of pursuing them I will simply highlight the possibly very troubling conclusion this suggests for the realist. This is that realist will have to endorse higher-order skepticism about her own normative knowledge: even if she can allow for the possibility that she knows normative facts, she will never be in a position to know this. Regardless of what world we are in, we won't be able to know whether it is a knowledge-friendly world. Thus while first-order skepticism is not a foregone conclusion for the realist, she will never be able to reassure herself (and us) that the normative facts are in fact just as epistemologically accessible as they are according to her anti-realist competitors.

References

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.

Dunaway, B. (Forthcoming), 'Luck: Evolutionary and Epistemic', *Episteme* .

Dunaway, B. (MS), 'The Metaphysical Conception of Realism'.

Dunaway, B. and Hawthorne, J. (Forthcoming), Scepticism, *in* W. J. Abraham and F. D. Aquino, eds, 'Oxford Handbook of the Epistemology of Theology', Oxford University Press.

Dunaway, B. and McPherson, T. (MS), 'Reference Magnetism as a Solution to the Normative Twin Earth Problem'.

Dworkin, R. (1996), 'Objectivity and Truth: You'd Better Believe It', *Philosophy and Public Affairs* **25**(2), 87–139.

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

Fine, K. (2001), 'The Question of Realism', *Philosophers' Imprint* **1**(1), 1–30.

Harman, G. (1986), 'Moral Explanations of Natural Facts–Can Moral Claims Be Tested against Moral Reality?', *Southern Journal of Philosophy* pp. 57–68.

Horgan, T. and Timmons, M. (1991), 'New Wave Moral Realism Meets Moral Twin Earth', *Journal of Philosophical Research* **16**, 447–465.

Horgan, T. and Timmons, M. (1992*a*), 'Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived', *Philosophical Papers* **21**(3), 153–175.

Horgan, T. and Timmons, M. (1992*b*), 'Troubles on Moral Twin Earth: Moral Queerness Revived', *Synthese* **92**(2), 221–260.

Kratzer, A. (1977), 'What "Must" and "Can" Must and Can Mean', *Linguistics and Philosophy* **1**, 337–355.

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.

Pritchard, D. (2004), *Epistemic Luck*, Oxford University Press.

Railton, P. (1986), 'Moral Realism', *The Philosophical Review* **95**(2), 163–207.

Schoenfield, M. (forthcoming), 'Moral Vagueness is Ontic Vagueness', *Ethics* .

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.

Street, S. (2008), Constructivism about Reasons, *in* R. Shafer-Landau, ed., 'Oxford Studies in Metaethics, vol. 3', Oxford University Press.

Wedgwood, R. (2007), *The Nature of Normativity*, Oxford University Press.

Williamson, T. (1994), *Vagueness*, Routledge.

Williamson, T. (2000), *Knowledge and Its Limits*, Oxford University Press.