Philosophical Perspectives, 18, Ethics, 2004

THE RETURN OF MORAL FICTIONALISM

Nadeem J. Z. Hussain
Stanford University

Fictionalism has made a comeback over the last two decades as one of the standard responses to ontologically problematic domains.[1] It has been applied to mathematics, modality, unobservables, identity claims, and existence claims.[2] Moral discourse has struck many as potentially ontologically problematic, but within contemporary analytic metaethics there has been no sustained defense of moral fictionalism.[3] Very recently moral fictionalism has also finally begun to return. On the dust cover of Richard Joyce's new book, *The Myth of Morality*—a sustained defense of moral fictionalism—David Lewis writes: "Moral fictionalism is an idea whose time has come."[4] In one sense, Lewis is right: in addition to Joyce's book there seems to be quite a bit of interest in moral fictionalism though much of it expressed merely in conversation or still in the publication pipeline.[5] The appeal of fictionalism often lies, I suspect, in its ability to look like a relatively less problematic alternative to both traditional noncognitivism and moral realism: we can do without the noncognitivist's problematic account of moral language and we can do without the realist's problematic metaphysics. In this paper I will not argue that moral fictionalism cannot work. Instead I will argue (i) that a correct understanding of the dialectical situation in contemporary metaethics shows that fictionalism is only an interesting new alternative if it can provide a new account of normative content: what is it that I am thinking or saying when I think or say that I ought to do something; and (ii) that fictionalism, qua fictionalism, does not provide us with any new resources for providing such an account.

1. Introduction

1.1 Metaethics: Some Methodological Comments

I will begin by expanding on the above claims before turning to a defense of them. First, the focus on normative content may seem excessively narrow and indeed it is. It was meant as shorthand for a more complicated view about the

task of metaethics. I will continue to use it as shorthand for the rest of the paper, but I will first explain the view that it is shorthand for.

Alex Miller suggests in his recent book, *An Introduction to Contemporary Metaethics*, that metaethics has broadened the range of questions it concerns itself with.[6] Philosophers used to think of metaethics as being "exclusively about language".[7] He illustrates this older view with a quotation from W. D. Hudson:

> [Metaethics] is not about what people ought to do. It is about what they are doing when they talk about what they ought to do.[8]

The difference now, according to Miller, is that metaethics is concerned with a range of questions: "what is the *semantic function* of moral discourse?…do moral facts (or properties) exist?… is there such a thing as moral knowledge?" What is the phenomenology of moral experience? What are the psychological structures involved in moral judgement and ensuing action? What kind of objectivity is available for moral judgements?[9] The focus has broadened, so the suggestion goes, from language to include metaphysics, epistemology, phenomenology and moral psychology.

Miller makes clear that "the various questions are not all independent", but I still think his way of describing contemporary metaethics does not emphasize sufficiently the unity of the field.[10] The aim of a metaethical theory, to modify slightly Hudson's definition, is to provide an account of what we are doing when we talk about what we ought to do and when we think about what we ought to do. Part of what we are doing when we talk and think this way, or at least what we do quite soon afterwards and plausibly as a result of talking and thinking this way, is acting in accord with—under the guidance of—these normative thoughts. So much is pretty uncontroversial. What I think this description of the aim of a metaethical theory brings to the fore though is that a theory does not count as a metaethical theory unless it has succeeded in explaining what is going on when we think that we ought to do something. In other words, if the theory fundamentally takes, even implicitly, the fact that we think normative thoughts as an explanans rather than an explanandum, then the theory has failed as a theory in metaethics. Neither is it sufficient simply to explain why I thought a particular normative thought.

It is this explanatory task that underlies and motivates the various questions in metaphysics, epistemology, phenomenology and moral psychology. Answering all of these questions is part of coming up with a thorough explanation of what we are doing. Hudson's description of the task thus remains in essence, I suggest, correct, even if perhaps we understand the nature of the explanation required somewhat differently than an 'ordinary language' philosopher would.[11]

Explanations come in many different kinds and grades, so it is hard to know how demanding to be here, but I think it is fair to say that the aim is to provide a comprehensive explanation that includes an account not only of why

we engage in normative thought, but also what it is to have that normative thought. It is hard to a priori rule out the possibility of doing one without the other: that is to explain what we are doing when we are thinking that we ought to do something without saying what it is to think that we ought to do something. Imagine the stubborn non-reductive realist. What are we doing when we think or talk about what we ought to do? Well, keeping track of what we ought to do and trying our best to do it. End, basically, of explanation. Some defensive comments may well be attached as addendum: there is no reason to think that there will be other language available to us to talk about either what it is to think that we ought to do something or what it is to be the case that we ought to do something than this very use of 'ought'. I do not think that the explanation is circular though it does rely on the recipient of the explanation being a competent participant in the practice. That said, such an explanatory account will be, other things equal, at a disadvantage relative to accounts that say more (the "other things equal" clause is of course a non-trivial requirement). Such an explanation—such a low grade of explanation—would only be accepted if we thought that we had no alternative.

Most of the competitors in contemporary metaethics are all accounts that attempt to provide a higher grade of explanation. This is true in particular of the account to which fictionalism can seem an attractive competitor, namely, traditional noncognitivism. It is also true, though this may be harder to see immediately, of the kind of realism relative to which fictionalism can seem to come at a cheaper ontological cost; a realism which we can accuse of paying a high ontological cost is one which goes beyond our stubborn non-reductive realist and provides some account of what it is for it to be the case that I ought to do something—some account that allows us to see what kind of ontological commitments we might be taking on in believing such a theory. A realism that *does not* provide such an account, for example our stubborn realist, has not said enough for us to claim that adopting such a view would indeed involve ontological costs relative to which fictionalism might be an improvement. We just do not have enough to go on to defend such a claim. Of course, we might well fall back on a low-grade explanation if we think that no high-grade explanation works.

My claim then that fictionalism is only an interesting new alternative to existing positions if it can provide a new account of normative content is really the claim that fictionalism is only interesting if it can provide a high-grade explanation, one that does try to say in detail what is we are doing when we talk and think about what we ought to do.[12] What this comes to will become clearer when we begin to see below more detailed comparisons of fictionalism with existing metaethical accounts. I will return at the very end to the question of whether a fictionalism that only provided a low-grade explanation might still be of use.

1.2 Kinds of Fictionalism

We do, however, need to distinguish between different kinds of fictionalism. We can draw a distinction between "hermeneutic" and "revolutionary"

fictionalism.[13] A hermeneutic fictionalist interprets the current discourse in fictionalist terms while the revolutionary fictionalist proposes fictionalism as a reform. Revolutionary fictionalism combines most naturally with an error theory about our existing discourse. Focussing on moral discourse, a revolutionary fictionalist would first argue for an error theory of moral discourse. She would claim, roughly, that moral sentences express propositions, that the existing moral practice involves believing these propositions, but that the propositions believed are systematically false. She would then go on to propose that we replace this practice with a fictionalist practice in which, rather than believing these propositions, we take up some other attitude towards them: we make-believe that *P* rather than believe that *P*.[14] In contrast the hermeneutical fictionalist would claim that, in some sense, we never really believed *P*, but that we have all along, in some sense, made-believe that *P*.

Comparing moral fictionalism to the metaethical positions already at the table is thus really a matter of two different kinds of comparisons depending upon whether one is considering hermeneutical or revolutionary fictionalism. Hermeneutical fictionalism competes relatively straightforwardly with other metaethical accounts: it claims to be a better account of our existing practice. The comparison of revolutionary fictionalism to some existing metaethical account, say norm-expressivism, involves two parts: (i) a comparison of the error-theory (that is a part of whatever revolutionary fictionalism is under consideration) to norm-expressivism as two competing accounts of our existing moral practice; (ii) a comparison of the two different proposals for how we should go on. The moral fictionalism will propose some kind of fictionalist practice of morality as a reform of our existence practice. Most other metaethical accounts, including norm-expressivism, propose, often implicitly, that we should go on with our existing moral practice despite, say, now coming to see norm-expressivism as the correct explanation of our practice.

It does make some sense to treat the error theory and the accompanying fictionalist reform as two parts of a single package—two parts of one theory that we want to call a form of revolutionary fictionalism. First, as I argue below, in order to explain such a pervasive error the error theorist will need to take us as having a deep need for something like morality. But this need will also make salient the question: what next? What to do once we have accepted the error theory? John Mackie's presentation of his error theory can appear quite puzzling in this regard.[15] Reading Mackie as a revolutionary fictionalist might make sense of what he thought he was doing when, after arguing for his error theory, he went on to develop what he called a "practical system of morality", but if so, the fictionalism remained quite implicit.[16] The revolutionary fictionalist provides fictionalism as an explicit response. The fictionalist's hope is that some of the resistance to adopting the error theory—resistance who's source lies in a practical concern with what will happen next or what to do next—can be overcome by presenting a fictionalist way of going on. The fictionalist, by combining error theory with a fictionalist replacement practice, is thus

supposedly presenting us with an under-appreciated alternative to the existing accounts and an alternative that goes beyond a mere resurrection of Mackie.

2. Revolutionary Moral Fictionalism

I will consider revolutionary moral fictionalism first. The attractions of a revolutionary moral fictionalism for a certain cast of mind are fairly straightforward. It is implausible that moral propositions are true because, in the spirit of Mackie, there is no objective prescriptivity, there are no intrinsically motivating objective properties, or no inescapable and overriding set of reasons, no "conclusive *to-be-done*ness", no objective authority.[17] The ontological and epistemological costs of being a moral realist may thus seem overwhelming. On the other hand noncognitivism has a hard time fitting with the apparent surface structure of our moral language and our intuitive sense that our existing moral practice does actually involve belief in moral facts. Responding to the Frege-Geach embedding problem requires the noncognitivist to develop a complicated account of moral language and thought—a cost when it comes to theory choice whether or not one thinks that the technical problems have been overcome. Such considerations can lead naturally to the thought that an error theory about our existing moral practice seems plausible. However, it is hard to imagine how we could simply go on without something very much like our existing moral practices. This would be hard to do though if we had to continue to believe what we knew were false claims. Fictionalism comes to the rescue. Just as I can make-believe that the tree stumps in the forest are bears while knowing that they are not, I can make-believe that murder is wrong while knowing that the claim 'Murder is wrong' is false.

The above crude sketch of the existing metaethical options, unfair in its brevity no doubt to those already mentioned, leaves out furthermore an important player, namely, naturalist realism.[18] Such realists give an account of the content of moral claims that is meant to remove the appearance of the ontologically problematic. The puzzle with such naturalist realist accounts is not ontological, but rather whether they have the right account of the content of moral claims. As we have seen already, the problem in metaethics is not deciding whether or not to pay the ontological and epistemological costs of moral realism—where the options otherwise are noncognitivism or error theory—but rather figuring out what the ontological cost of realism is in the first place. What makes this hard is that we do not have a compelling account of normative content. Revolutionary moral fictionalism, when deployed as fictionalism usually is in other domains and as it has been so far in metaethics, does not represent an under-appreciated alternative to the metaethical positions already at the table because it fails in the end to deal with what is in fact the real problem for metaethics, namely, the problem of normative content. Revolutionary fictionalism in the standard deployment has neither a new way through the

problem of normative content nor does it provide a way around the problem of normative content—or so I shall argue.

I will start by sketching the odd dialectical space that the revolutionary fictionalist finds herself in. This sketch will again perforce be very crude, but I hope just accurate enough to give us a sense of the problems that various alternative strategies of dealing with this dialectical space would face and the risks of succumbing to certain temptations in avoiding these problems. In §3 I'll turn to a detailed consideration of one contemporary version of revolutionary moral fictionalism to show how it succumbs to some of these temptations—understandably enough as I hope this section will show.

2.1 Error Theory

Start with the revolutionary fictionalist's error theory. An intuitive picture of moral discourse is that moral judgements involve ascribing moral properties to things (persons, actions, states of affairs). Now the error theorist thinks that the best account of our moral practices is indeed that these practices involve ascriptions of moral properties. So far the error theorist agrees with a realist in being a cognitivist about moral discourse. The error theorist goes on however to argue that such ascriptions are systematically false or unwarranted.

How does the error theorist go about defending these claims? The error theorist has to argue that moral properties are such that nothing instantiates them or that our very notion of these properties is such that it makes no sense to think that these properties do exist or perhaps even could exist. Take some purported moral property, say, moral rightness. One straightforward way to argue that all claims of the form *x* is right are false is to say that our notion of being right, or of rightness, is such that no *x* is, or could be, right.[19] The error theorist thus has to provide an account of moral language and thought that can compete with the accounts provided by competing theories in metaethics.[20] Thus the very first dimension along which revolutionary fictionalism would be assessed is whether the account of moral language that its error theoretic part provides can stand up to the alternatives.

In order to distinguish herself from the noncognitivist, the error theorist needs to claim that moral language does have truth conditions. Note that the error theorist and the realist thus share the task of vindicating the surface cognitivity of moral discourse. So one option for the error theorist is simply to join the moral realists in their attempts to come up with truth conditions for moral discourse. There are of course several different realist projects. For some projects it is a bit hard to see how they could be of much help to the error theorist. Take Cornell-style non-reductive naturalist realism. Such accounts defend their realism by claiming that irreducible moral properties play an ineliminable role in some of our best explanations. The error theorists would hardly want to use this as a starting point for giving an account of the content of moral claims since they could hardly want moral properties to play such a role.

Similar problems arise with attempts to use the semantics of natural kinds or causal theories of reference.[21] An error theorist will not want the burden of denying that there may well be some naturalistic property, or natural kind, that moral language is in some sense "tracking". The error theorist though will deny that it makes sense to see ourselves as talking about this naturalistic property. Consider the analogy with talk of witches. There may well be some naturalistic property—women who challenge local social hierarchies—that talk of witches tracks. But an error theorist about witch talk would want to deny that these women are witches. Why? Well, because they do not really live up to our notion of witches: they do not fly on broomsticks, turn people into frogs, etc. Analogously the error theorist about moral discourse will want to claim that given our moral concepts, even if there is a naturalistic property or kind that our use of moral terms tracks, this isn't sufficient grounds for claiming that our moral claims are true in virtue of this property being instantiated.

The error theorist is thus pushed to what seems very much like the business of providing an *analysis* of moral concepts. She could just try to give a non-circular conceptual analysis of moral terms in terms of non-moral concepts. We know the applicability conditions of these non-moral terms, so we can see if anything in the world satisfies these non-moral concepts and thus we can see if anything satisfies the moral concept. The error theorist will defend the view, basically, that nothing does satisfy moral concepts.

The problem is that such analyses have often seemed quite implausible. Moore's open-question argument may not pick out some specific naturalistic fallacy; however, there remains a widespread sense that most such attempts at analysis miss something essential about moral concepts. It is indeed the very normativity that seems to go missing. The worry usually spreads to analyses in non-normative terms of normative vocabulary in general—at least practical normative vocabulary. This is what motivates in part noncognitivism and various kinds of non-naturalist realisms. Now the existence of such widespread misgivings is not a decisive argument against any particular proposed analysis nor an argument against proposing new analyses. What does follow though is that given the current dialectical situation in metaethics, a view that rests essentially on its success at providing a noncircular conceptual analysis of moral concepts will only be an interesting alternative to existing positions if it has something more compelling, at least equally compelling but new, to bring to the table than the existing proposals for such an analysis.[22]

But why should a revolutionary fictionalist have to provide such an analysis? Forcing a revolutionary fictionalist into providing a compelling analysis of moral concepts better than those on offer from realists may seem quite perverse. After all, part of what might attract one to an error theory about the moral in the first place is the thought that there is something deeply mysterious about moral concepts and the moral properties they supposedly pick out. Morality, one thinks, is an ideology and mystification is the life-blood of ideologies. Surely it would be no surprise, then, if some fundamental unclarity is essential to

morality's ideological role. Given this essential unclarity, no surprise, then, if moral concepts seem to systematically escape analysis. Fictionalism shouldn't try to go through the normative content problem, but around it.

I agree that this is the more natural way for a moral fictionalist to go and this is the path that, as I shall try to show below, Joyce actually takes despite initial appearances. However, before we consider the details of his view, we can already see that there are costs involved in taking such an approach to defending an error theory. The costs are of two kinds: (i) it is harder to argue that moral properties don't exist or aren't instantiated and, most importantly, (ii) the revolutionary fictionalist will have a hard time explaining both how false moral beliefs play whatever is their sustaining functional role in existing moral practice and how moral make-belief will be able to play some similar role in the proposed fictionalist replacement practice.

The first cost may not seem so high. We might think that there are two fairly straightforward ways of showing that moral concepts aren't satisfied without giving a full analysis. To start with, we might give a partial analysis. Perhaps we can articulate some necessary condition or conditions such that we know, *a priori* or *a posteriori*, that nothing could or does satisfy this condition or these conditions. Who knows what all the notion of God involves, but perhaps the notion of omnipotence taken together with some reflections on the possibility of unmovable objects, convinces us that no such thing can be. The second approach, compatible with the first, involves arguing that the best explanatory account of *how* we come to have our moral beliefs makes it unlikely that these processes of belief acquisition are epistemically respectable. Our revolutionary fictionalist could then, at least, argue for systematic epistemic error.[23] A realization that there is such error should lead us to give up our moral beliefs and perhaps leave us open to the fictionalist's proposed moral make-belief.

The normative content problem, though, remains. The problem, for the revolutionary fictionalist, has two parts. First, the revolutionary fictionalist has to defend his error theory against a straightforward objection—an inverse, moral version, if you like, of what David Lewis called "the incredulous stare" objection.[24] How could such a massive, systematic error that pervades our lives simply have gone more or less unnoticed for so long? And how could such an error continue to keep us in its grips despite the possibility of massive moral error having been emphatically brought to our attention by the likes of Nietzsche and Marx?

The error theorist answers this objection by claiming that the beliefs, though false, play an important role in our lives. It is this important role that then explains their staying power. Indeed, as we've seen suggested already, the mystificatory nature of moral beliefs is an essential part of the explanation of the role of the beliefs in our lives and their ability to sustain themselves.

The content problem however is this. It is hard to see how one could provide a story of the role of such beliefs in our lives without explicating the *content* of these beliefs since it must be in virtue of their distinctive content that

these beliefs play the distinctive role they do. Finding some necessary conditions of *a*'s being *F* such that we have sufficient grounds to argue that no claim of the form *a* is *F* could be true, will not provide us with a sufficient account of what it is to believe that *a* is *F* such that we could argue that believing that *a* is *F* plays some central role in our lives—a central enough role that systematic error would escape our notice.

Imagine that we have a partial analysis such that

$$\forall(x)F(x) \rightarrow G(x)$$

And so perhaps if I believe that *a* is *F* then I believe that *a* is *G* and *a* is *X* where *X* is, so to speak, the part of the content of the belief that the error theorist is leaving unanalyzed. Now if the full explanation of the role of moral beliefs in our lives could be carried out merely by considering the content *a* is *G*, then that would undermine the claim that there is indeed any further content to the beliefs beyond just the belief that *a* is *G*. The conclusion one will be pressed on to draw then is that one has in fact given a full analysis of moral concepts.

On the other hand it is hard to see how one could provide a compelling explanation in the moral case—and in the normative case in general—of the role of beliefs of the form *a* is *F* (where *F* is a normative predicate) in our lives where one is not willing to spell out the content of "*F*" beyond stating that a necessary condition of something being *F* is that it is *G*. What do I mean by "compelling"? This is a placeholder to indicate the kind of explanation one needs given the dialectical situation the defender of error theory for morality is in. Take as an example a simple version of the error theory often ascribed to Mackie according to which it is motivational internalism that generates the problems. Motivational internalism of the relevant kind is of course a controversial thesis. But let's grant for the sake of argument that if true, it would show that moral claims are false. Now this error theorist faces the "incredulous stare objection" and needs to explain how such false beliefs could be so widespread. Imagine that the error theorist simply says that believing that murder is wrong tends to prevent people from hurting others in their society and societies that have such beliefs are the ones that survive and get to propagate their beliefs.[25] No further story is told about the content of such beliefs in the explanation or the details of how such a belief leads to action. The problem in the metaethical case is that for many of us—no doubt those with Humean sympathies of various kinds—it is simply a puzzle how normative beliefs by themselves could lead to action. A theory that simply claims that they do without any further explanation is at a dialectical disadvantage. One can perhaps avoid developing a theory that explicates the content of normative beliefs by positing a *de dicto* desire to do the right thing or simply insisting that Humean worries are misplaced and that normative beliefs can directly lead to action. However, such claims have often seemed implausible to many if not most participants in the debate. So the question won't be whether the error theorist has some explanation,

but rather whether they have an explanation that can compete with the other proposed explanations.

Part of what we have to remember here is that the error theorist must inevitably treat the playing field as *not* being level. The error theorist in the moral case is asking a lot of us. We have to reject a massive part of our original belief structure. The error theorist, if we accept his theory, can show us why it makes sense to reject all those beliefs. But inevitably when considering whether or not to accept the error theory, we assess it against the beliefs we have prior to such acceptance, and then accepting the error theory appears to be epistemically hazardous. Matters would be simple if one could have a deductive argument for error theory from indubitable premises, but such is rarely the case. The partial analysis on the basis of which the claim that there is an error is defended is never simply uncontroversial. If the explanatory side of the error theory can't hold its own in the metaethical debate of how moral beliefs lead to action, then it is in trouble. And I suggest that it won't be able to do this without giving a full account of the content of moral beliefs and an account of how, given this distinctive content, moral beliefs play a central and distinctive role in our lives—a role such that we would continue to have these beliefs despite their being systematically false.

2.1.1 Objection: Hard to Detect Errors

What I've said so far may seem unfair to the error theorist. I claimed that if the error theorist is merely providing something like a noncircular conceptual analysis, then she hasn't brought anything new to the table unless her analysis is more compelling—or at least equally compelling but new—than the existing proposals for such analyses. But surely what is distinctive in what the error theorist is bringing to the table is the argument for *error*. For example, as Richard Joyce to some degree will—details and caveats in §3—we could take on, say, Michael Smith's analysis of moral terms but argue that in fact our moral claims are false because there is no convergence among rational agents: "there is no *single* thing that fully rational creatures would all want us to do in the various circumstances in which we find ourselves."[26] The distinctive contribution of the error theorist is precisely the argument, whatever it is, in favor of this claim of error—here the argument for non-convergence. Finally the error here is one that might well be quite hard to detect and so perhaps no "incredulous stare" is warranted. Thus no explanation of how this error could go undetected is needed, and so no explanation of the persistence of these beliefs in terms of the part of the content that is error-laden, so to speak, is needed.[27]

Responding to this objection requires a more careful discussion of what the adjective "systematic" comes to in the claim that an error theory in metaethics posits a systematic error. I have many moral disagreements with my fellow humans on a vast range of moral issues. I tend to think of them as committing mistakes—as being in error—when they claim that abortion should be made

illegal, that my moral obligations to my fellow citizens are different than my moral obligations to foreigners, that retributive punishments are morally justified, and so on. Indeed there are wide swathes of practical normative judgment that I tend to think are mistaken: I find assessments of moral responsibility systematically troubling because I find myself worried about the conceptions of agency that lie behind them and I intuitively find it hard to buy any judgments involving notions of honor though I would be harder pressed to articulate why. Of course I'm not exceptional in this way. What is important to note though is that in most of these cases I and others would, if pressed, deploy first-order normative arguments. It is not as though we think that moral properties just are not or cannot be instantiated. It is just that some people—perhaps lots of them—are wrong about which things have which moral properties.

Consider an analogy with witch talk. One can well imagine the epistemically, and perhaps religiously, sane true-believer in witches unwilling to rely on *judicium Dei* to determine who a witch is.[28] Such a person would think that many of the judgments of others about witches are, in one important sense, systematically mistaken: most of their claims about who are witches are probably false and certainly not warranted. Our true believer though still thinks that there *are* witches and might well be quite concerned that the wrong detection techniques are being used. Contrast his case with the standard contemporary view that witches don't exist. Such a view is systematic but in a different sense. The worry is more metaphysical. Being a witch involves the power to do magic and magic doesn't exist. This doesn't have to be the claim that there is no possible world where there is magic, but that in our world there is no magic and so no witches. The grounds for thinking there are no witches is a suspicion about the very property of being a witch—at least of some of its essential components. The systematicity of this view is intuitively of a different order than that of our epistemically sane, but nonetheless, true-believer in witches.[29]

Returning to metaethics then, I don't get to be an error theorist just because I think many of my fellow humans are quite mistaken in their moral views. The metaethical error theorist is someone who thinks that moral properties are dubious on very general grounds. Whether I think retributive punishments are wrong *or* right, there is a mistake involved in my judgment just in virtue of my using moral concepts. But this kind of error, even if subtle—even if hard to detect—is then so pervasive that, I suggest, we would be unconvinced by any argument for this error unless that error itself played some important functional role in the practice.

Return to the example of an error theoretic use of Smith's view. We now see that the claim wouldn't just be that convergence often doesn't occur or that others are wrong about when convergence occurs. Rather the view would be that convergence never occurs. It is hard to consider the matter in such abstraction of course. We'll look later at Joyce's version of such an argument. However, I think it is already hard to believe that we would be willing to accept a theory that both claimed that we were committed to convergence in our moral claims

and that such convergence never occurred, if we weren't also given some story for why it would make sense for us to be so committed to convergence. Such a story could show, for example, the functional role in the practice of morality of a commitment to convergence. The error results from the distinctive content of normative claims—their use of normative predicates—and so this functional story will have to show that this error-generating content plays a crucial role in the practice.[30]

2.2 Fictionalism as Reform

I said that the normative content problem for the revolutionary fictionalist has two parts. The first, as we've just seen, is that without specifying the content of moral beliefs the revolutionary fictionalist will have a hard time constructing the kind of compelling explanation of our current practices needed to sustain the error theory against the "incredulous stare" objection. The second problem emerges in his proposed fictionalist replacement.

Arguably many of the participants in the metaethical debate implicitly accept an error theory in a "soft" sense: they grant that there is no completely comprehensive account of our existing ethical practices and discourse that would not end up taking as mistaken or confused *some* of the commitments that are part of the practices.[31] Their aim is to reform our ethical practices in order to remove these confusions; they provide us with a new ethical practice partly constituted by an accurate conception or theory of this new practice. They argue that the proposed practice would be relevantly similar to the original practice in that it serves the essential point and purpose of the original practice.[32] Given such a project of developing a reformed practice, it is precisely the errors that an error theorist posits that one might well think would be the first things to be removed in the transition to a new practice and to the accurate conception of itself that is part of this new practice.

The revolutionary fictionalist has to argue against such a reforming move. She wants to convince us that we should adopt a moral fictionalism. The heart of her claim presumably has to be that in fact the proposed reforms won't be able to serve the point and purpose of the original practice. Let's put aside the stranger possibility of a noncognitivist reformer for a second just to be able to articulate the point at issue here more cleanly. A realist reformer will claim that we replace a moral practice in which we believed things of the form 'murder is wrong' with a practice in which technically we are replacing the predicate 'wrong' with some close neighbour, say, 'wrong*'. The reformer thinks we should keep using the old term 'wrong', but we should now understand it as expressing wrong*. Though 'murder is wrong' was, literally construed, perhaps false, 'murder is wrong*' is true and so we can believe murder is wrong*. The fictionalist has to argue that pretending that murder is wrong is more effective than believing that murder is wrong*. Now again the fictionalist faces the normative content problem. Without precisely specifying the content of what

it is to think that murder is wrong it is hard to see how the fictionalist could make the case in favor of pretending that murder is wrong as opposed to believing that murder is wrong*.

The revolutionary fictionalist thus faces a difficult choice. On the one hand, providing a full noncircular conceptual analysis of moral concepts in non-moral terms seems to involve a daunting task—a task only the bravest take on and a task at which even the bravest, one might think, have failed. This approach, as I've granted, also may not quite seem to be in the right spirit. After all it is the mysteriousness of purported moral properties and facts that tempts one to error theory in the first place. On the other hand, not providing a full analysis—trying to go around the normative content problem—creates serious difficulties. Without a full account of the content of moral beliefs it is hard to see how, first, the fictionalist can articulate the kind of explanations she needs in order to defend her error theory and how, second, she can articulate the proposed pretence and defend its preferability over reformist realist alternatives.

The discussion so far has been very abstract. I will turn in the next section to a discussion of a concrete example of revolutionary moral fictionalism. I'll focus on the moral fictionalism defended in Richard Joyce's recent book, *The Myth of Morality*. However, I think it is important to have a sense in more abstract terms of the topology of the dialectical space in which the revolutionary moral fictionalist has to operate. This will allow us to see more easily why Joyce's project ends up looking the way it does and why it faces the problems I'll raise for it below.

3. A Case Study: *The Myth of Morality*

Richard Joyce, in his fascinating and elegantly argued book, *The Myth of Morality*, argues for a revolutionary moral fictionalism and so his account comes in two parts: the error theory and the proposal of a fictionalist replacement. Let's consider the error theory part first. Officially, Joyce takes the route of providing a noncircular conceptual analysis. Joyce's strategy in arguing for his error theory is to accept reasons internalism for moral claims. He calls this Mackie's Platitude:

> MP: It is necessary and a priori that, for any agent x, if x ought to ϕ, then x has a reason to ϕ.[33]

He combines this with an analysis of practical reasons similar to those proposed by Bernard Williams and Michael Smith:

> S has a subjective reason to ϕ if and only if S is justified in believing that S+ (S granted full information and idealized powers of reflection) would want S to ϕ.[34]

He argues though that Smith is wrong to think that fully informed, fully rational agents will converge in their judgements and so, Joyce thinks, we have to accept the relativity of reasons. What reasons S has does depend on what desires S has since what S+ wants is still a function of S's desires—full information and full rationality don't allow S+ to rise above, so to speak, the contingency of S's desires. Moral claims on the other hand purport to present us with oughts that are quite independent of our desires—the oughts of morality are categorical. The argument for the error theory is then as follows:

1. If x morally ought to ϕ, then x ought to ϕ regardless of what his desires and interests are.
2. If x morally out to ϕ, then x has a reason for ϕing.
3. Therefore, if x morally out to ϕ, then x can have a reason for ϕing regardless of what his desires and interests are.
4. But there is no sense to be made of such reasons.
5. Therefore, x is never under a moral obligation.[35]

The relativity of reasons of course is what lies behind (4).

Recall our earlier discussion of the dialectical situation faced by the revolutionary fictionalist. Officially then Joyce has chosen the option of providing a noncircular conceptual analysis of moral terms, indeed of practical reasons in general. Understood in this manner, when it comes to providing an account of the content of our moral claims, it seems as though Joyce has nothing more to offer than Michael Smith. What is different is the denial that indeed fully informed rational agents will converge in their desires—a point already raised by others and a possibility recognized by Smith.[36]

However, the worry about Joyce's view isn't just that it doesn't provide a compelling new alternative to existing accounts of our moral discourse. Though he officially accepts Smith's analysis of moral concepts, he, as I shall argue below, ends up taking the more natural route for a moral fictionalist, namely, that of seeing moral concepts as fundamentally mysterious and unanalysable. This, as suggested in the previous section, is completely understandable but leads to problems—problems that I think Joyce doesn't overcome.

As we have just seen, Joyce deploys the Smith analysis to argue for an error theory. He thus faces the incredulous stare objection. As I've already pointed out, an error theorist owes us an explanation of the presence and persistence of such a systematic error. Joyce recognizes this. He says:

> A proponent of error theory—especially when the error is being attributed to a common, familiar way of talking—owes us an account of why we have been led to commit such a fundamental, systematic mistake. In the case of morality, I believe, the answer is simple: natural selection.[37]

Here though, as I shall argue below, is the point at which the normative content problem rears its head again despite the official account of content having already been given.

What is Joyce's evolutionary story? He points to the evolutionary advantages of being concerned to help family members and being able to cooperate with others. What mechanisms would then be selected for? Joyce considers for example the possibility that "we would be granted by evolution a general *desire* to look after the interests of our family members."[38] However, he says:

> The problem with desires is that they are unreliable things.…Desires can be overridden by stronger, contrary desires, and desires for intense, immediate gratification can be temporarily re-evaluated to the detriment of satisfying longer-term, calmer desires.…Because of these kinds of limitations of desire, an individual does better (in the sense of being more reproductively fit) if she has her desires in favor of family members supplemented by a sense of *requirement* to favor family members. It is not merely that an individual *wants* to help out his sister's son, but he feels that he *ought* to—he feels that he *must*.[39]

Recall that Joyce is trying to argue that we would be selected to have false *beliefs* about the moral. Now our normal belief-forming mechanisms are also presumably the product of natural selection for him, but they involve a certain sensitivity to evidence—sensitivity enough for us to be willing to call them beliefs. The purported systematically false moral beliefs would have to be systematically isolated, so to speak, from the dispositions to respond to evidence. Joyce's theory thus acquires two related burdens: one conceptual and one explanatory. The conceptual burden is that given that the mental attitudes in question seem not to be systematically responsive to evidence regarding the truth of their contents, it isn't clear that we have sufficient reason to call them beliefs in the first place.[40]

The explanatory burden is harder to pin on Joyce, but the idea roughly is this. *A priori* there are more options for natural selection. Even if we grant Joyce, as he claims, that the "very flexibility and adaptability of desires in many areas of life … is their great virtue" there are other possible, and perhaps actual, mental states to choose from. Why not standing, overriding intentions to help others constituted by norms quite different from desires such that they do have the requisite stability?[41] Evolution doesn't have to pick optimum solutions, but, intuitively, there does seem to be a cost to using beliefs since those beliefs would have to be set up so that the normal responses to evidence do not occur.[42]

But let's put these worries aside. I find the thought that this, to use Joyce's phrase, "sense of requirement" could be selected for *prima facie* plausible. However, I'm relying of course on my own intuitive sense of the sense of requirement that goes with thoughts that I ought to do something. The question though is whether Joyce's official account of the content of thoughts of the form I ought to ϕ fits his explanation for the error.

Recall that, according to Joyce, to think that I ought to ϕ is just for me to believe that were I fully informed and fully rational I would want myself to ϕ. Now how is this belief supposed to provide the requisite sense of requirement, or, to put the point more carefully, how is this belief supposed to do the job that, for example, a desire to help others couldn't?

Consider the kind of case that the sense of requirement is supposed to help with:

> After a hard day's hunting and gathering, one may be so weary as to have one's desire to look after the children enervated or quite dissipated. The long-term satisfactions of child-rearing may be under-appreciated due to phenomena like distraction, weakness of will, or simple exhaustion. The desire is likewise under threat if one's nephew, say, has accidentally broken one's favorite spear, thus provoking intense annoyance.[43]

Now imagine that at this point I recall that I also believe that my fully informed, fully rational self would want me to look after the child (or my nephew). Would that make any difference? Why should it? Why should that belief make any more difference than the belief that my child wants me to look after him? Notice that the fact that my full-informed self's having of this belief is supposed not to be a function of my current set of desires doesn't seem to make any particular difference either. It would make a difference if I were already wired somehow to be moved by thoughts of what my fully rational self would want.[44] Alright, then, one might think, evolution will select for a strong desire to do what my fully rational self would want me to do. But, as we've seen Joyce already argue, desires are fickle, overridable creatures. What would presumably do the job is for me to have a sense that I am *required* to do what my full rational self would want me to do. But what is this sense that I am required to ϕ if it isn't the belief that I want myself to ϕ if I were fully informed and deliberating correctly?

One might think that Joyce has provided an answer already in his defense of the claim that questioning practical rationality itself is incoherent.[45] I think it is useful to see why that argument won't help here.[46] Here's the argument:

> To ask "Why should I ϕ?"…is to imply that one is in the business of accepting reasons, that one is able…to recognize reasons…that one is disposed to participate in deliberative activity, that one values such things as evidence and truth.…If this is correct, then we are in a position to see that Jill *does* take the desires of Jill+ as reasons, for those desires just are what Jill would desire for herself if she were fully reflective and epistemically successful. In other words, the question "I recognize that if I were to deliberate properly on the matter, armed with all and only relevant true beliefs, I would desire my actual self to ϕ, but what is that to me?" is not something we need take seriously, for just in asking the question one would be demonstrating one's valuing of deliberation and truth.[47]

It may be that one cannot ask, "Why should I ϕ?" without thus expressing a commitment to accepting reasons. However we are currently trying to see if Joyce can construct an argument to show why it would be evolutionarily beneficial to be in this business in the first place. Our hunter gatherer may not be able to sensibly ask such a question either before or after getting into the business, but we can still ask how giving him moral beliefs is supposed to help him do better. We want to know how the commitments involved in being in this business are embodied in the agent and what difference they make.

Joyce's explanation of the widespread error of morality thus looks incomplete. The error we make in morality, according to Joyce, is to believe that I morally ought to help my child. That is to believe that I ought to help my child regardless of my desires and interests. That in turn is to believe that I have a reason to help my child regardless of my desires and interests. That in turn is to believe that my fully-informed, fully-rational self would want me to help my child regardless of my desires and interests. Joyce wants to explain why we have such generally false beliefs by pointing to their evolutionary usefulness. Having this belief would supposedly make it at least more likely that I would help my child. However, it does no such thing without positing in addition something that will get me to go along with my full-informed self. By his own lights a desire will not do. We need a sense of requirement—I need to feel that I ought to do what my fully informed self would want me to do. But what is this feeling? What is this ought?

Joyce comes close to seeing the problem here but in the end, I suggest, doesn't quite see it, as some revealing comments show. He is officially committed to a Humean view of motivation and so not surprisingly he writes, "I am not claiming that a belief in a categorical requirement *alone* can provide motivation, for that would be to retract the Humean view of motivation that was argued for" earlier.[48] What he says next is, I think, revealing:

> John Stuart Mill commented that there is a "disposition to believe that a person who sees in moral obligation a transcendental fact, an objective reality belonging to the province of 'Things in themselves,' is likely to be more obedient to it than one who believes it to be entirely subjective, having its seat in the human consciousness only." Mill goes on to disagree with this supposition, claiming that whether we think of the sanction of morality as internal or external…, a person's obedience to the moral law depends on her "subjective feeling." There's a sense in which Mill is quite correct: mere awareness of the authoritative basis of a moral claim will not suffice to carry motivation with it.…What determines whether morality makes any difference to us is the attitude we take towards it, not its ontology. But Mill fails to recognize that the attitude we ultimately take towards morality can be influenced by how "external or internal" we conceive morality's authority to be. Whether I obey a person's orders depends on whether I want to, of course—but if I believe that person to have legitimate authority over me then it is far more likely that the desire to comply will be prompted. Similarly, if I believe that a kind of action *must* be done, then the

> entertainment of other courses of action may be excluded from my deliberations, affecting my motivations. This is the distinctive behaviour-modifying value of moral concepts: they are imbued with maximal authority.[49]

But this leaves the explanation incomplete. What is it to believe in "legitimate authority"? How *are* concepts "imbued with maximal authority"? The real metaethical work is still to be done. The normative content problem has not been solved in a way that allows him to deal with the "incredulous state" objection.

Let me be clear. This is not to deny the plausibility of the claim, to use Joyce's words, that a sense of requirement, a feeling that he ought to ϕ, or the feeling that he must ϕ would indeed, allow an individual to do evolutionarily better. My claim is that the account of the thought that S morally ought to ϕ in terms of believing that S+ would want S to ϕ doesn't fit into the explanatory story. If we rely on our intuitive sense of what Joyce is pointing to when he talks of a sense of requirement, the explanation might seem to make sense, but as long as we don't have an analysis, we don't really have a compelling explanation in the sense discussed in §1.

I think that Joyce must at some level be aware of this. In his extended discussion of morality and evolution where he lays out his explanation for why we commit the "fundamental, systematic mistake" of believing moral claims, the analysis of moral concepts in terms of reasons and reasons in terms of desires under full information and full rationality does not play much of a role. Instead Joyce repeatedly uses different phrases that express our vague sense that moral facts or properties purport to have some mysterious authority or normativity. Here's a list of expressions he uses in addition to the sense of requirement and the feeling that one must do certain things already mentioned: moral concepts are "imbued with maximal authority" (140), "certain rules of conduct have intrinsic authority, that mustn't be questioned, that are there in the nature of things" (141), "authoritative 'must-not-be-doneness'" (168), "the quasi-mystical commitments embodied by moral discourse" (170), and later, in trying to asses why a "moral framework" is so important to us, he suggests:

> Perhaps…the appeal to an authority that lies outside our chosen ends, that cannot be articulated but has a powerful 'felt' quality, may satisfy some need for an essentially mysterious element underpinning human affairs.[50]

Talk of fully-informed desires seems not to play much of a role here.

Joyce may well be right. Somehow it is this mysterious element, this authoritative 'must-not-be-doneness' that is so hard to articulate that does indeed play a crucial role in getting us to take care of children when tired and keep our promises even when the possibilities of defection sing their sweet siren songs.[51] The problem is that without something like an analysis of normativity the explanations deployed in defense of the error theory appeal only to these vague metaphors—metaphors that will strike a chord with some but hardly seem to do the job of rigorous explanation needed to answer the "incredulous

stare" objection. What we need is a careful and precise account of the mysterious normative element in the content of moral beliefs and an explanation of how this content both allows these beliefs to play the proposed role in our psychologies—the role of providing a sense of requirement or must-be-doneness—and allows us to claim that these beliefs are false.

3.1 Moral Oughts, Non-Moral Oughts and Practical Rationality

The normative content problem comes out even more clearly once we take a closer look at the distinction in Joyce's work between moral oughts and non-moral oughts. Joyce, as we saw above, grants that there are normative reasons. I do have a reason to ϕ as long as my fully-informed, fully-rational self would want me to ϕ. No error here. Moral oughts involve an error because moral oughts claim that I have a reason quite independent of my desires and interests. Moral claims purport thus to be, in this sense, inescapable. Whether they give me reasons is supposed to be independent of my desires. This for Joyce is the official "queerness" of the moral.[52] And it is this queerness that supposedly provides a compelling argument from queerness for a moral error theory. Such claims won't be true because my fully-informed, fully-rational self has the desires he or she does only because I have certain desires and not others.

Now, it would be natural to take Joyce's talk of authoritative must-be-doneness in his evolutionary explanation of our error in believing moral propositions as equivalent to the inescapability feature of moral claims. I've already argued that inescapability, understood in terms of the fact that S+ desires S to do certain things quite independent of what S desires, won't by itself explain why an S that believes he morally ought to help her child will help her child in the face of temptation or tiredness. When we look closely at many of Joyce's claims of the role of thoughts of authoritative must-be-doneness in evolution, we come to see something interesting. Authoritative must-be-doneness doesn't seem just to be restricted to moral claims. And authoritative must-be-doneness turns out not to be what's special about moral claims.

Joyce as I've just said does not hold an error theory about practical reasons and practical rationality in general. So when I pursue my best interests I am acting on reasons that I really do have. But even in such cases, the "mere fact that I justifiably believe ϕing to be in my best interests simply does not ensure that I will ϕ" (184). Joyce claims:

> Moral thinking…functions to bolster self-control against such practical irrationality. If a person believes ϕing to be required by an authority from which she cannot escape, if she imbues it with a "must-be-doneness," if she believes that in not ϕing she will not merely frustrate herself, but will become reprehensible and deserving of disapprobation—then she is more likely to perform the action. In this manner, moral beliefs can help us to act in an instrumentally rational manner.[53]

But practical irrationality occurs not just in cases that we normally think of as involving morality. This comes out clearly in one of Joyce's other examples:

> Suppose I am determined to exercise regularly, after a lifetime of lethargy, but find myself succumbing to temptation. An effective strategy will be for me to lay down a strong and authoritative rule: I must do (say) fifty sit-ups every day.[54]

Now Joyce just doesn't say this, but if we are convinced by his evolutionary story that I believe propositions of the form 'I morally ought to ϕ because this will help me keep my promises, not steal, etc. despite temptation', then don't we have just as much reason to think that the role of thoughts of the form 'I ought (non-morally) to ϕ' is also to keep me from succumbing to short-term temptations and keep me on track in the pursuit of my long-term interests? Surely by Joyce's own lights, a thought that I ought to do fifty sit-ups plays a very similar role to the thought that I morally ought not to steal. Such thoughts are instrumentally effective in helping me pursue my long-term self-interest (or at least those of my genes). But then it doesn't make sense to see authoritative must-be-doneness as restricted just to *moral* oughts. Both intuitively, and given the evolutionary story, authoritative must-be-doneness turns up for non-moral oughts too.[55]

But now Joyce has a problem. Officially authoritative must-be-doneness is explained in terms of inescapability. And inescapability is explained in terms of a supposedly special feature of moral oughts, namely, that whether I morally ought to ϕ is supposed to be quite independent of my desires and interests. This is precisely why moral claims are supposedly false, but it is precisely the absence of this independence from desires and interests that accounts for why normal oughts aren't systematically false.[56] That is why normal oughts aren't, in this sense, inescapable. But then inescapability can't explain the authoritative must-be-doneness for non-moral oughts. And this in turn I think should make us suspicious of the official account of authoritative must-be-doneness for moral oughts in terms of inescapability.[57]

Furthermore, if we look more carefully at parts of Joyce's evolutionary story, I think it becomes clear that inescapability, understood as always in Joyce's particular sense, isn't really the difference between moral oughts and other practical oughts anyway. This comes out in his discussion of the development of a conscience. Here, he claims:

> Guilt is the negative feeling an agent has when she believes that she has acted in a way she ought not have, and it is a necessary and important element of any system of values deserving the name "moral".[58]

Guilt is to be clearly distinguished from *regret*, the feeling one has when one suffers straightforward practical irrationality—when one fails, for example, "to act so as best to satisfy a desire".[59] This may well be a fruitful way to distinguish the moral from the non-moral but notice that inescapability again does nothing

to help clarify this distinction. Why should guilt be associated with inescapability? How do we identify the feeling of guilt without indeed identifying it in terms of the appropriate emotional response to a particular kind of normative claim?

3.2 Joyce's Fictionalist Reform

Let me turn finally to show what difficulties the normative content problem raises for the second part of Joyce's revolutionary fictionalism, namely, his proposal that we adopt morality as a fiction. Again, the goal here is to illustrate in a concrete instance the dialectical difficulties raised in §1 for the revolutionary fictionalist strategy in general.

The first problem emerges from the fact that the revolutionary fictionalist wants to claim that we ought to adopt moral fictionalism now that we realize that our moral beliefs are false. If the ought in this claim was the moral ought, then the fictionalist would be uttering what is by his own lights a falsehood. Joyce's response is to say that the ought here is merely an ought of instrumental reason and not a moral ought: "It is just a straightforward, common-or-garden, *practical* 'ought'".[60] What is essential to this strategy then is to prevent any taint of the error ascribed to moral oughts from spreading to practical oughts. The discussion in the previous section though should make us nervous precisely about this. What is mysterious about moral oughts is their normativity, their authoritative to-be-doneness. This feature might well be a reason to think there is really no such authority in the world and a reason to think that evolution would have made us think that there is. But it is just as plausible to think that evolution would have selected for thinking there is authoritative to-be-doneness in the domain of non-moral oughts. No surprise then perhaps that many feel that reductive accounts of non-moral oughts leave something out too. However, for the revolutionary fictionalist non-moral oughts threaten to become as mysterious and problematic as moral oughts. Such non-moral oughts would then no longer be a safe place from which to make the claim that we ought to accept revolutionary fictionalism.

The second difficulty for the revolutionary fictionalist emerges when he tries to argue in favor of preferring his proposed pretence over reformist realist alternatives. A concrete example will help here. Let's take Peter Railton's moral realism not as an accurate account of our current moral practices but as a proposed replacement. Railton's account is a form of naturalist realism. To simplify drastically, an action's being right just is a matter of the action's maximizing non-moral good where the non-moral good of each is counted equally. ϕ-ing is non-morally good for S if and only if S+ (fully informed and fully *instrumentally* rational) would want S to want ϕ. The account of non-moral good thus should be acceptable to Joyce.[61] There is no claim of convergence here. What is non-morally good for two agents may well be a function of their different desires. Railton's account of the truth-conditions for moral claims doesn't, unlike Smith's, depend on convergence under full information and

rationality. This allows for the possibility that the action that might be most instrumentally rational for achieving an agent's non-moral good might not be the morally right action to do.[62] But for Railton this doesn't at all undermine the truth of the moral claim.

It follows perhaps that Railton is not committed to reasons internalism. And perhaps that is sufficient reason to think that he hasn't provided a correct account of our existing moral practice.[63] However, here we're only considering Railton's view as a proposed reforming realist alternative to fictionalism. Our fictional Railton might well grant that reasons internalism, what Joyce calls Mackie's platitude, perhaps just was the error in our existing moral practices. Once we drop that requirement, accept Railton's account of the truth-conditions of moral claims (technically perhaps moral* claims), we can go on with our new moral practice that involves believing that murder is wrong (technically perhaps wrong*) rather than make-believing that murder is wrong. There is no reason why these claims should be systematically false since presumably there could be facts about what would maximize overall non-moral good. Indeed murder is plausibly just the kind of thing that will not maximize overall non-moral good.

So why accept moral fictionalism rather than such a moral reformist realism? Here's Joyce's own argument for his moral fictionalism. Moral beliefs, Joyce argues, get us to cooperate with each other. Enlightened self-interest also in general recommends such cooperation; however,

> the instrumental value of moral beliefs lies in their combating of weakness of will, their blocking of the temporary revaluing of outcomes that is characteristic of short-sighted rationalizations, their silencing of certain kinds of calculation. And it is my contention that these desiderata can be satisfied, to some extent, even if the moral claims are not believed.[64]

In particular he thinks that fictionalism will do better than two alternatives: Abolitionism: "where moral discourse has been abandoned altogether" and second, "where moral claims are (somehow) believed, despite the fact that evidence of the falsity of such beliefs are glaring".[65] I won't consider his argument against these two alternatives. I'll focus instead on the alternative he leaves out, namely, moral reformist realism.

Joyce has to argue that make-believing that murder is wrong will result in more cooperation and coordination than believing murder is wrong*. Notice that the choice then is not between, on the one hand, having no belief like murder is wrong or continuing to believe something that one knows is false, and, on the other hand, make-believing that murder is wrong. Rather the choice is between make-believing that murder is wrong and believing that murder is wrong*. What the fictionalist again intuitively needs to say is that believing that murder is wrong* doesn't have authoritative must-be-doneness. Now it is important to remember that make-believing murder is wrong doesn't really have authoritative must-be-doneness either. So the real issue is whether pretending

that something has authoritative must-be-doneness will be more or less effective than believing that murder is wrong*. If all there was to pretending that there is authoritative must-be-doneness is for S to pretend that S+ would want her to ϕ, it is hard to see why this would be particularly effective. Why would this make-believe have more motivational impact than the real belief that genuine non-moral good would not be maximized? Or in the case of murder would actually be decreased? Again without a clear and compelling account of authoritative must-be-doneness, of normativity, it is hard to see how the revolutionary fictionalist can make his case here.

4. Hermeneutic Moral Fictionalism

Let us now turn briefly to hermeneutic moral fictionalism. There are many different ways one could develop hermeneutic moral fictionalism and different kinds of problems that might arise.[66] Some of the worries raised for revolutionary moral fictionalism in the preceding sections will apply fairly straightforwardly to many of the ways in which one could develop hermeneutic moral fictionalism. I will try to focus in this section on whether a hermeneutic fictionalist approach might provide resources or suggestions for developing a fictionalist account that go beyond those already considered. I will also restrict the scope of the discussion by taking as my target natural ways in which hermeneutic moral fictionalism could deploy Stephen Yablo's recent work.

Yablo uses fictionalism in his ongoing project to defuse worries about the existence of abstract objects. Indeed he introduces a rationale for his hermeneutic fictionalism that can seem quite attractive in the wake of our discussion of the problems facing revolutionary fictionalism:

> At one time the rationale for fictionalism was obvious. We had, or thought we had, good philosophical arguments to show that X's did not exist, or could not be known about if they did. X's were *obnoxious*, so we had to find an interpretation of our talk that did not leave us committed to them.
>
> That form of argument is dead and gone, it seems to me. It requires very strong premises about the sort of entity that can be known about, or that can plausibly exist....
>
> But there is another possible rationale for fictionalism. Just maybe, it gives the most plausible account of the practice. It is not that X's are intolerable, but that when we examine X-language in a calm and unprejudiced way, it turns out to have a whole lot in common with language that is fictional on its face. If one now asks which elements of everyday speech are fictional on their face, the answer is the figurative elements.[67]

The figurative element that Yablo focuses on is metaphor. Metaphor has various features that make it attractive for Yablo's project and for the project of providing a hermeneutic fictionalism. To start with we use metaphor extensively.

Consider one of Yablo's examples: "Jimi's on fire" spoken, say, of an athlete playing well. Jimi is not literally on fire. We do not have to worry about whether he will recover from his burns. It is natural also to say that, in some sense, we are pretending that he is on fire. There is a game of make-believe in which, when he is playing well, we get to say that he is on fire.[68] There is no ontological problem raised by the puzzling fact that Jimi can both be on fire and not in pain. And so on. A hermeneutic fictionalism seems appropriate when it comes to the claim that Jimi is on fire.

Using these features of metaphor, and a theory of metaphor as involving make-believe, Yablo claims that we can give a hermeneutic fictionalist account of mathematics. Talk of numbers is figurative. This explains why we do not have to worry if there is an abstract object referred to when we say that the number of planets is nine. Nonetheless such language is useful. Just as we can convey information about how well Jimi is doing on the field by saying that he is on fire, we can convey facts about equinumerosity by talking as if there were certain abstract objects.[69]

This kind of fictionalism—a fictionalism that appeals to metaphor—has a number of features that might appeal to a moral fictionalist. Let us take a look at a couple.[70] First, we apparently do not have to ascribe any error to the practitioners so we do not need to explain why this error would have gone unnoticed. It is not as though we were ever in any way deluded about Jimi's being on fire: we do not sigh with relief at Jimi's fate now that we have learned the truth.

Second, the "'pregnancy' of metaphors" could perhaps be used to account for failures of analysis of normative concepts.[71] If the use of normative terms is somehow metaphorical, then no surprise perhaps that analysis fails. Paraphrases of metaphors have an open-endedness to them—the phrase "and so on" often haunts the paraphrases of, at least, most metaphors. This open-endedness is perhaps what lies behind the open-question argument faced by analyses of normative concepts in other terms. Consider a proposed analysis as follows:

"ϕ-ing is morally right" means the same thing as "ϕ-ing maximizes utility"

A speaker who asks, "Yes, I know ϕ-ing maximizes utility, but is it morally right?" is expressing no more conceptual confusion than a speaker who says, "Yes, I know Juliet is the warmth of my world, that my day begins with her, but I would not say she is the sun."[72] He might be reluctant to call Juliet the sun, despite sharing many of Romeo's feelings, because for him thinking of her as the sun would be to think of her as cruel and distant, beating down upon him mercilessly. Perhaps a different cultural, or should I say climactic, context might make this more natural: Mirza Ghalib describes his beloved's face as the midday sun to express the traditional cruelty and mercilessness of the beloved of the *ghazal*.[73]

Our fictionalist might claim that an analysis of a moral concept is, despite its official aims, in effect a paraphrase of a metaphor. But the nature of the relationship between a metaphor and its paraphrase, even where the paraphrase is adequate for an occasion of utterance, allows for a different speaker to either question the adequacy of the paraphrase or refuse the invitation to use the metaphor in the way recommended by the paraphrase—or perhaps to insist on using the term used in the metaphor in a different metaphor. All this without linguistic confusion.

Such a fictionalism might be able to explain other otherwise puzzling features of moral discourse. Supervenience can perhaps be explained by the "principle of generation" of the relevant game of make-believe, to use Kendall Walton's terminology.[74] Principles of generation allow real facts to generate fictional truths. We can use this to explain why the fictional truths of morality supervene on non-moral truths. Finally, perhaps, we can hope to explain the distinctive motivational and affective impact of normative vocabulary by pointing to the distinctive impact of figurative language.

There is much then that is attractive to the appeal to metaphor though obviously we do not yet have a theory. The problem for a metaphor-based hermeneutic fictionalism in metaethics is that it needs to provide a detailed account of what we are doing when we say or think that, for example, killing little children for the fun of it is wrong. To say that there is metaphor here is not yet to give an account of what we are doing when we utter such a sentence or have such a thought. What role does such a thought play? What does it do and how does it do it?

Again, I will begin by looking at Yablo's attempts to answer these kinds of questions when he gives us a hermeneutic fictionalism for abstract objects, and see if this provides a way for hermeneutic moral fictionalism to answer such questions. He engagingly puts it this way:

> Suppose we discuss the alleged theft of the platonic objects the way we would any other alleged crime. Means, motive, opportunity—are all these elements present?
>
> The question of means is: how would a job like this be pulled off, where objects appear to be in play but really aren't? The question of motive is: why would anyone *want* to fabricate these objects in the first place? The question of opportunity is: how could a job this big be pulled off without anyone noticing?[75]

The answer to the question of means is to begin with, as we have already seen, metaphor: "'numbers' are conjured up as metaphorical measures of cardinality."[76] This leads us to motive. Yablo gives the example of "*the average so-and-so*" (295). He writes:

> When someone says that
>
> (S) The average star has 2.4 planets,

> she is not entirely serious; she is pretending to describe an (extraordinary) entity called "the average star" as a way of really talking about what the (ordinary) stars are like on average. True, this *particular* metaphor can be paraphrased away, as follows:
>
> (T) The number of planets divided by the number of stars is 2.4.
>
> But the numbers in T are from an intuitive perspective just as remote from the cosmologist's intended subject matter as the average star in S.…The nominalist therefore proposes that *T* is metaphorical too; it provides us with access to a content more literally expressed by
>
> (U) There are 12 planets and 5 stars or 24 planets and 10 stars or…[77]

But, Yablo suggests, (U) is either not allowed by English because infinitely long sentences are not allowed, or, though allowed, is much less preferable than (S). (S) is to be preferred "for its easier visualizability", "its greater suggestiveness ('that makes me wonder how many moons the average planet has'), the way it lends itself to comparison with other data ('the average planet has six times as many moons as the average star has planets'), and so on".[78]

Thus at the heart of Yablo's account is the idea that such figurative language can convey information about the real world. Yablo emphasizes that metaphors can be "essential" in various ways; the work they do sometimes cannot be done by some literal alternative.[79] Nonetheless our motive is to convey information about real objects and it turns out that these metaphors of abstract objects give us a means to do so.

The moral hermeneutic fictionalist needs both a motive for moral language and an account of how moral language understood as metaphor could provide adequate means to this end. Again there are many possibilities for how such a theory could be developed, but I think we can see some of the difficulties that lie in the path of developing any plausible hermeneutic metaphor-based moral fictionalism by considering the obvious motives one might propose. In the spirit of Joyce, we might think that thinking that something is right is a way of getting myself, or others, to stay the course in the face of various desires that might pull me in various directions. Normative thought is essential to avoiding being a wanton.

How though might moral language and thought achieve this end? Recall that in Yablo's account there is a distinction between the literal content and the metaphorical content:

> A metaphor has in addition to its *literal* content—given by the conditions under which it is true and to that extent belief-worthy—a *metaphorical* content given by the conditions under which it is 'fictional' or pretence-worthy in the relevant game.[80]

David Hills puts the point in terms of a distinction between the "presented thought…entertained in a spirit of assertion" and the "presenting thought… entertained in a spirit of *pretence*."[81] As Hills emphasizes, though,

> any particular way of *taking* the words metaphorically is built on a specific assignment of pre-metaphorical senses to the words in question. To take words metaphorically is to assign content to them twice over. In the special but standard case of a routine freestanding metaphorical assertion, we take the same set of words to express two different thoughts: a presenting thought and a presented thought.[82]

The obvious question to ask then is what the presented and the presenting thought are in the case of moral fictionalism.

Notice that we need to say something about both in order to give a coherent story of motive and means. In the case of Yablo's account for mathematics, for example, the presented thought is a thought about equinumerosity and the presenting thought is a thought about abstract objects. What are the equivalent thoughts in the case of moral fictionalism?

This demand may seem unfair. After all was not the point about the pregnancy of metaphor and its paraphrase precisely that we could not give a complete statement of the presented thought? The "and so on" seemed often ineliminable. The defender of fictionalism needs to be careful at this point though. We are not asking for a paraphrase that can be stated without the "and so on". Rather we need enough of a paraphrase to understand enough of the presented thought to see why presenting such a thought would help achieve the purported end of the practice. Such an answer is compatible with the presence of an "and so on" added on at the end of the paraphrase. Refusing to give any account of the presented thought would leave us without any explanation of how the metaphorical practice is a plausible means given the purported motive. Notice that we would also be without an explanation if it were explicitly claimed that the part of the presented thought that achieved the end was present only in the part referred to by the "and so on".

We can now see how the issues here are closely related to the problems raised for revolutionary fictionalism. If the hermeneutic fictionalism simply claims that the thought presented is that I ought to ϕ, then we will have two concerns. First, what is this thought? Two, what remains of the fictionalism? Why should we take moral language as figurative?

An obvious suggestion though for the presented thought is that it is a thought about the presence of "right-making" properties.[83] As Stephen Darwall puts it, "no particular thing can be *barely*…right".[84] It is in virtue of the right-making properties that the action has the property of rightness. This, it seems, is true of normative properties in general. These right-making properties could be non-normative properties. We thus might be able to claim that moral language presents thoughts about non-moral facts: it involves asserting the presence of the relevant right-making properties. So to say that Jill's killing of John was wrong might convey the information that Jill's killing of John had at least one wrong-making property. An example of such a property might be that the action was done intentionally with the aim of satisfying her desire to inherit John's money. Call this property *F*.

But why not just say that ϕ-ing is F? The hermeneutic fictionalist has a number of at least initially plausible things to say here. First, as in the case of Yablo's example of the "the average so-and-so", saying that ϕ-ing is wrong might be an easy way of expressing a content whose literal expression, as in the case of (U) above, would require an ugly, (perhaps infinitely) long, disjunction. ϕ-ing has property F *or* ϕ was done accidentally but Jill could easily have avoided killing John if she had been paying attention to the traffic signs *or* John was not an enemy combatant *or*....

Second, again following a suggestion of Yablo's, we may need the metaphor for its "presentational force".[85] Just saying that ϕ-ing is F would miss out on the practical nature of normative language. Surely, says the fictionalist, in saying that ϕ-ing is wrong I, normally, at least, want to get you not to ϕ and keep myself from ϕ-ing. The thought has to play the appropriate role in rationalizing action and so on. The presentational force, we can hope, is what will allow us to capture the practical role of such thoughts.

We can perhaps see how the presenting thought, that ϕ-ing is wrong, might be useful in presenting the ugly disjunction; there is a pretence that there is a single property that all of these cases instantiate, but it is the question of "presentational force" here that is the tricky part and the location of the problem of normative content for a metaphor-based hermeneutic fictionalism. If we just say that the pretence is that there is a normative property—that the presenting thought is the thought that ϕ-ing is wrong—then we are left with our metaethical task unfinished since what we wanted to know was what it is to have such a thought.

It helps to see how hard the problem of normative content here is going to be by thinking through a couple of attempts to respond to it. The list of attempts will not exhaust logical space but I think it does point out how hard it will be to find something that will work in this logical space—and perhaps also a better sense of what we are looking for in this logical space. Imagine for a second pretending that there is objective prescriptivity, to use Mackie's phrase, in the following sense: we pretend that the action itself, or the fact constituted by the action and the right-making properties, addresses us literally—literally within the fiction—by commands. "Do me!" says the action. Or so we pretend. Would that do the job?

Not really, I think. This might be part of what we have to pretend, but the question is whether we have a full account of the kind of pretence required by moral language understood as metaphor. One way to bring out the fact that we do not really have an explanation here is to see that an explicit pretence to this effect does not seem at all like thinking the relevant thought and to notice what precisely seems missing, namely, normativity. I can pretend that an object is addressing me with a command to ϕ, but not think at all that I ought to ϕ. What is missing from the pretence is precisely the thought that I *ought* to obey the object. In other words, within the game of make-believe itself I need to think a normative thought, but then the game does not really explain what it is to think a normative thought.

Now the kind of pretence involved, according to theories like Hills' and Yablo's, in metaphor is not as conscious or not as self-conscious, for lack of better words, as the kind of pretence I was imagining in the previous paragraph. We might well think that it is precisely the absence of this kind of self-consciousness that allows the figurative language of rightness and wrongness to do its work. However, it is not good enough, I think, merely to say that. One has to show how this lack of self-consciousness makes the requisite difference. I find it hard to see how one would do that.

Or imagine that I am just pretending that the action has some property called 'wrongness', but that I also am pretending that I desire to do actions with this property. We are adding the desire in the hope that this can capture "presentational force". The hope, as with many naturalist moral realisms and noncognitivisms, is that the motivational states of the agent—actual or counterfactual—can provide the necessary materials for explaining the phenomenology of normativity. Think of the term 'wrongness' as just a label we give this property so pretended. Add on a bit more to the pretence if you like: we can perhaps assume that the right principle of generation is in place such that the ugly disjunction gets presented. What is essential to the pretence is that presentational force is supposed to be accounted for by the desire to do actions with this property. To bring out that such a pretence does not do the job either, imagine changing the label to 'ghalat'. The problem again is that I can pretend that an action is ghalat and pretend that I desire to do actions with the property ghalat, without thinking that the action is wrong—without thinking that I ought not to do the action.

This is not to deny that pretending that I desire something may well get me to do some action. Consider a somewhat different example. I can pretend that these clumps of grass are broccoli and I can pretend that I desire broccoli (in my case this would *have* to be pretence). As part of my game, then, I might well, really, bring the clumps of grass towards my mouth and, pretend no doubt, to put them in my mouth. All this might be part of the pretence. Again though, as this example perhaps even more clearly shows, there is no reason to think that engaging in this pretence would be at all like thinking that I ought to eat grass.

The pretence *as a whole* needs plausibly to seem like thinking that I ought to ϕ; however, it is hard to see how it could even come close to that unless it is part of the specification of the game of make-believe that I pretend that I ought to ϕ—that within the pretence I am to think that I ought to ϕ—but then the make-believe story will not explain what it is to think that I ought to ϕ.

The hope will remain that some redescription of the pretence will make it plausible to think that in pretending that P I am thinking ϕ-ing is wrong. Perhaps if I pretend that I want myself to want whatever has F, or that I would so want under full information, and so on. One will be tempted to try applications of all the other existing metaethical accounts. Perhaps one of these, or some variation, will turn out to work. I cannot claim to have pointed to some "pretence fallacy" that shows that all such attempts will fail. As I have

mentioned already, the history of accounts that claim that thinking that I ought to ϕ, or that ϕ-ing is right, just is thinking that ϕ-ing is F where F is some non-normative property does not inspire much confidence. The hermeneutical fictionalist does have an additional move to make. She attempts to explain thinking that I ought to ϕ in terms of *pretending* that ϕ-ing is F. There is no reason, though, as far as I can see, that this additional move will meet with any greater success as long as F is not a normative property.

5. Conclusion

Fictionalism is presented as a new option in metaethics. It is presented as a forgotten alternative to moral realism or noncognitivism. If you cannot get yourself to believe in moral facts, don't go noncognitivist, just become a fictionalist. However the hard task in metaethics is coming up with an account of normativity—or rather what it is to have thoughts with normative content thus allowing for the possibility that there might be no such thing as normativity. That is the real puzzle. The situation in moral fictionalism is perhaps different from many other domains to which we apply fictionalism. Start with the simple Santa Claus case. I do have some notion of what Santa Claus is supposed to be like and so I think I can both confidently claim that Santa Claus does not exist and that I know what it is to pretend that Santa Claus will deliver presents on Christmas Eve. Move now to a slightly harder case: a kind of fictionalism about possible worlds. I have some sense of what it would be to be some kind of realist about possible worlds and to hold with David Lewis "that our world is but one world among many."[86] Maybe I do not really, but it is perhaps easy to think I do because I have some sense of what the actual world is like. So again I have a fairly clear sense of why I find it so hard to be a realist about possible worlds and I have, I think, a sense of the pretence involved in some kind of modal fictionalism and, more importantly in the present context, what such a pretence would do for me and how.

When it comes to metaethics though the real challenge is providing a compelling account of what it is to even think normative thoughts and how those thoughts manage to shape our practical lives. Perhaps one of the cognitivist or noncognitivist accounts on the table will convince us at the end of the day. But this is where the real debate is: what is it to have thoughts with normative content and how do such thoughts lead to action? A fictionalist position will not really have much to add to the debate unless it has something new to add to the existing accounts of normative content.

One could perhaps always just say, "Look, I'm a competent user of moral language. I have a grasp of moral concepts. I can't give you an analysis, but I can tell you that there isn't anything in the world that lives up to them. And given my grasp of the concept, I just know that the other vindicatory metaethical accounts on the table are wrong. And so their accounts of the moral psychology are wrong too." A revolutionary fictionalist of this stripe might go

on to insist, "I don't know the explanatory details that show how a belief that something is wrong leads one not to do it, but it happens everyday. Fortunately pretending that something is wrong does to. And I'm just going to continue on with a simulacrum of this practice where I pretend that things have moral properties." A hermeneutical fictionalist of this stripe would continue the initial stubborn thought differently: "Since nothing in the world lives up to these concepts, I think it is clear that we're just pretending that they're satisfied. I can't really say anything more about what's involved in this pretence than just that it is pretending that things are wrong (or using the word "wrong" metaphorically). But just because I can't say anything more doesn't mean it isn't true."

Nothing I have said in previous sections shows, I think, that such positions are incoherent or mistaken. My discussion focussed on the dialectical situation of the fictionalist where she sees her task as convincing others and as taking on competing positions in contemporary metaethics. That's where things of course get hard.[87]

Notes

1. Cf. Gideon Rosen, "What is Constructive Empiricism?," *Philosophical Studies* 74, no. 2 (1994): 168. Fictionalism of various kinds was a standard nineteenth-century strategy for dealing with the ontologically problematic. Relevant names would include, I would argue, David Friedrich Strauss, Feuerbach, Lange, Bentham, Nietzsche, Mach, Avenarius and of course (edging into the twentieth century) Vaihinger.
2. See, for example, Hartry H. Field, *Science without Numbers: a Defence of Nominalism* (Princeton: Princeton University Press, 1980); Hartry H. Field, *Realism, Mathematics, and Modality* (Oxford and New York: Blackwell, 1989); Arthur Fine, "Fictionalism," *Midwest Studies* 18 (1993); Bas C. Van Fraassen, *The Scientific Image* (Oxford: Clarendon Press, 1980); Gideon Rosen, "Modal Fictionalism," *Mind* 99 (1990); Mark Crimmins, "Hesperus and Phosphorus: Sense, Pretense, and Reference," *Philosophical Review* 107, no. 1 (1998); Gareth Evans, *The Varieties of Reference* (Oxford: Clarendon Press, 1982); Kendall Walton, "Existence as Metaphor?," in *Empty Names, Fiction, and the Puzzles of Non-Existence*, ed. Anthony J. Everett and Thomas Hofweber (Stanford, Calif.: CSLI Publications, 2000); and Stephen Yablo, "A Paradox of Existence," in *Empty Names, Fiction, and the Puzzles of Non-Existence*, ed. Anthony J. Everett and Thomas Hofweber (Stanford, Calif.: CSLI Publications, 2000).
3. John Mackie's account, as I briefly suggest below, is not explicitly presented as a form of fictionalism. A fictionalist interpretation may nonetheless be a charitable reconstruction. See his J. L. Mackie, *Ethics: Inventing Right and Wrong* (London: Penguin, 1977).
4. Richard Joyce, *The Myth of Morality* (New York: Cambridge University Press, 2001). See also Richard Joyce, "Moral Fictionalism," in *Fictionalist Approaches to Metaphysics*, ed. Mark E. Kalderon (Oxford: Oxford University Press, forthcoming).

5. Work in the pipeline includes Daniel Nolan, Greg Restall, and Caroline West, "Moral Fictionalism" (MS, 2002) and a book by Mark Kalderon called *Moral Fictionalism*. Assessing whether the worries raised in this paper do actually apply to these works will have to await their final versions and publication of course.
6. Alexander Miller, *An Introduction to Contemporary Metaethics* (Cambridge: Polity Press, 2003).
7. Miller, *Contemporary Metaethics*, 3.
8. W. D. Hudson, *Modern Moral Philosophy* (Garden City: Anchor Books, 1970), 1. Quoted at Miller, *Contemporary Metaethics*, 3.
9. Miller, *Contemporary Metaethics*, 2.
10. Miller, *Contemporary Metaethics*, 2.
11. The association of Hudson with 'ordinary language' philosophy is Miller's (Miller, *Contemporary Metaethics*, 3).
12. We will also want to make sense of how some thoughts can apparently be normative for us without having normative content and how various thoughts can stand in normative relations without having normative content.
13. See John P. Burgess and Gideon A. Rosen, *A Subject with No Object: Strategies for Nominalistic Interpretation of Mathematics* (Oxford and New York: Clarendon Press; Oxford University Press, 1997), 6–7 and Jason Stanley, "Hermeneutic Fictionalism," *Midwest Studies in Philosophy* 25 (2001): 36. See also John P. Burgess, "Why I Am Not a Nominalist," *Notre Dame Journal of Formal Logic* 24 (1983).
14. "Make-believe" is a placeholder for a range of attitudes, and a range of contents for those attitudes, which different versions of fictionalism might propose.
15. Mackie, *Ethics: Inventing Right and Wrong*. Mackie himself seemed to think that his error theory did not have to undermine our first order moral views (16, 105–106).
16. Mackie, *Ethics: Inventing Right and Wrong*, 169. One can interpret Mackie as a reforming realist. The general possibility of being a reforming realist is something we will return to below. For suggestions that we should read Mackie as a fictionalist, see Richard Garner, "Are Convenient Fictions Harmful to Your Health?," *Philosophy East & West* 43, no. 1 (1993).
17. The quote phrase is from Allan Gibbard, *Thinking How to Live* (Cambridge: Harvard University Press, 2003), 5. Gibbard is not talking about morality narrowly construed. More on this below.
18. It leaves out two important players actually. The other is the Kantian. Trying to figure out how contemporary Kantian accounts of morality fit into the traditional metaethical typology is a difficult task in its own right and one I will leave for another occasion.
19. For the purposes of this paper, I will from now on ignore the question of whether moral properties could be instantiated in other possible worlds even if they are not in ours.
20. As we shall see, it can also take over and defend an existing account, but argue that given this account, moral claims are false.
21. See Richard N. Boyd, "How to Be a Moral Realist," in *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord (Ithaca and London: Cornell University Press, 1988) and also Peter Railton, "Moral Realism," in *Moral Discourse and*

Practice: Some Philosophical Approaches, ed. Stephen Darwall, Allan Gibbard, and Peter Railton (New York and Oxford: Oxford University Press, 1997; originally published *The Philosophical Review* 95 (1986): 163–207) and Peter Railton, "Naturalism and Prescriptivity," *Social Philosophy and Policy* 7, no. 1 (1990).

22. See Michael Smith, *The Moral Problem*, ed. John McDowell, Philip Pettit, and Crispin Wright, *Philosophical Theory* (Oxford: Blackwell, 1994), Michael Smith, "Internalism's Wheel," *Ratio* 8, no. 3 (1995), and Frank Jackson, *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Oxford and New York: Clarendon Press, 1998).
23. Cf. Joyce, *Myth*, 158–74.
24. David K. Lewis, *Counterfactuals* (Cambridge: Harvard University Press, 1973), 86 and David K. Lewis, *On the Plurality of Worlds* (Oxford and New York: Blackwell, 1986), 133.
25. My thanks to Ariela Tubert for the example and for pressing me on this point.
26. The quote is from Smith, *Moral Problem*, 201. Smith does not think there is any reason to be sceptical now even if this is a logical possibility (201). As I will note again below, it is important to bear in mind that Smith's analyses are what he calls a "summary-style, non-reductive analyses" though they are still meant to square moral talk with naturalism. Moral properties will still turn out to be natural properties even if "a thoroughly explicit and reductive network-style analysis of our moral concepts in naturalistic terms" is not a viable option (57–58). I don't think this additional complexity affects the point being made in the main body of the text.
27. My thanks to Stewart Cohen and Robert Mabrito for pushing this objection.
28. I am sure there were many, but I have to plead ignorance of historical reality.
29. Perhaps I should say "relatively epistemically sane".
30. The following objection might be raised to my insistence that an explanation of the functional role of false moral beliefs requires giving an account of the normative content of these beliefs. A conceptual-role semantics might claim though that "the content of concepts are determined by their functional role in a person's psychology" (Gilbert Harman, "(Non-Solipsistic) Conceptual Role Semantics," in *Reasoning, Meaning, and Mind* (Oxford and New York: Clarendon Press; Oxford University Press, 1999), 206). One worry then is that I'm getting things exactly the wrong way around since functional role determines content. A version of such a view has been developed for metaethics in Ralph Wedgwood, "Conceptual-Role Semantics for Moral Terms," *Philosophical Review* 110, no. 1 (2001). Wedgwood argues that "the meaning of the moral term 'better than' is entirely determined by" conceptual role (20) and that this meaning "must determine which property (if any) the term stands for; that is, the meaning of the term must determine a certain *condition*, such that the term stands for a property if, and only if, the property meets that condition. Mackie's error theory must claim that there is no property that meets that condition" (3). Thus, for all I've said, there is another strategy available for arguing for an error theory that has not been directly addressed.

    For reasons of space, I just will not be able to address this purported alternative approach adequately in this paper. I find myself in strong agreement

with the arguments against Wedgwood presented in Laura Schroeter and Francois Schroeter, "A Slim Semantics for Thin Moral Terms?," *Australasian Journal of Philosophy* 81, no. 2 (2003): 191. Indeed, I think that these arguments are related in significant ways to the problems I am suggesting here for error theory and fictionalism. Nonetheless, a careful assessment of conceptual role semantics for moral terms, and an assessment of possible arguments for error theory or fictionalism based on such a semantics, will have to await another occasion.

31. See, for example, the notion of "reforming definitions" in Richard B. Brandt, *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979), 5–11, C. L. Stevenson, "The Emotive Meaning of Ethical Terms," *Mind* 46 (1937) and the discussion in Stephen Darwall, Allan Gibbard, and Peter Railton, "Toward *Fin de siècle* Ethics: Some Trends," *The Philosophical Review* 101, no. 1 (1992): 127, 45–46, 78, and 85–86.
32. Claims to the effect that such and such is the point and purpose of our ethical practices are often hard not to hear as claims about what *should* be the point and purpose and of our ethical practices. We can then well wonder how we are to take these claims: are they to be taken as claims made within the old practice or made within the new practice? I won't consider here how reforming metaethical theories deal with such puzzles. We will however consider below how moral fictionalism deal with similar puzzles about the status of the "ought" in their claims that we ought to take up the fictionalist practices.
33. Joyce, *Myth*, 38. Joyce wants to resist calling this reasons internalism "so as not unduly to multiply 'internal/external' distinctions" (38 n. 10). But this particular multiplication is already widespread and I think it is more confusing at this point to deny, or not to make clear, that (MP) is what most people call reasons internalism.
34. I am simplifying somewhat but this is the analysis we can draw out from comments at various places. See Joyce, *Myth*, 75, 83, 100. I'm also basically ignoring his distinction between subjective and objective reasons. It's unclear whether Williams provides an analysis (see Bernard Williams, "Internal and External Reasons," in *Moral Luck: Philosophical Papers 1973–1980* (Cambridge: Cambridge University Press, 1981), Bernard Williams, "Internal Reasons and the Obscurity of Blame," *Logos* 10 (1989), and Bernard Williams, "Replies," in *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, ed. J.E.J. Altham and Ross Harrison (Cambridge and New York: Cambridge University Press, 1995)) and one should remember that Smith is providing a "summary-style, non-reductive" analysis (Smith, *Moral Problem*, 56–58). I do think that some of the details of Smith's view would matter if one were trying to defend him against Joyce's criticisms but I won't take up that task here. Peter Railton's view looks similar, but it isn't a straightforward analysis so I think it is a mistake for Joyce to lump it in with Smith's view (Joyce, *Myth*, 75 and Railton, "Moral Realism,"). I discuss Railton's view further below though for different purposes.
35. Joyce, *Myth*, 77.
36. For example, David Sobel, "Do the Desires of Rational Agents Converge?," *Analysis* 59, no. 3 (1999).
37. Joyce, *Myth*, 135.

38. Joyce, *Myth*, 136.
39. Joyce, *Myth*, 136–37. Emphases in the original.
40. A hermeneutical fictionalist could attempt to take advantage of this. The attitude is not one of belief but in some way is already that of make-belief.
41. Consider the range of motivational attitudes—constituted by different norms of stability—that Bratman has argued for over the years. See Michael E. Bratman, *Intention, Plans, and Practical Reason* (Cambridge: Harvard University Press, 1987) and the articles in Michael E. Bratman, *Faces of Intention: Selected Essays on Intention and Agency*, ed. Ernest Sosa, *Cambridge Studies in Philosophy* (Cambridge: Cambridge University Press, 1999).
42. Cf. R. Jay Wallace, review of Joyce, Richard, *The Myth of Morality*, Cambridge, 2002., *Notre Dame Philosophical Reviews* (2003).
43. Joyce, *Myth*, 136.
44. In a sense this is what Smith claims. If I'm rational, then I'll be moved by such thoughts just because being rational involves caring about what my rational self would want. The problem for Joyce though is that he needs to explain why we are "wired" this way.
45. Joyce, *Myth*, 49–51 and 82–83.
46. I'm not suggesting that Joyce would want to use it at this point.
47. Joyce, *Myth*, 83.
48. Joyce, *Myth*, 140.
49. Joyce, *Myth*, 140.
50. Joyce, *Myth*, 228.
51. Joyce, *Myth*, 184.
52. Joyce thus takes himself to be agreeing with Garner (Joyce, *Myth*, 31 and Richard Garner, "On the Genuine Queerness of Moral Properties and Facts," *Australasian Journal of Philosophy* 68 (1990)). I don't think Garner is committed to an analysis in terms of the non-normative though.
53. Joyce, *Myth*, 184.
54. Joyce, *Myth*, 215. As the rest of the discussion of the example in his text suggests, Joyce is thinking of the "must" as being part of a descriptive claim despite his talk of an authoritative rule here. I must do fifty sit-ups in the sense that that is what is necessary in order for me to be fit. I suspect that part of the appeal of the example turns on the ambiguity between the reading of "must" as descriptive or normative, but the rest of my argument does not rely on a normative reading.
55. Cf. Christine Korsgaard, "The Normativity of Instrumental Reason," in *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut (Oxford: Oxford University Press, 1997); "Skepticism about Practical Reason," in *Moral Discourse and Practice: Some Philosophical Approaches*, ed. Stephen Darwall, Allan Gibbard, and Peter Railton (New York and Oxford: Oxford University Press, 1997).
56. Of course, any particular non-moral ought claim could be false for other reasons.
57. Inescapability understood in the precise sense that Joyce understands it.
58. Joyce, *Myth*, 137–38. It's true that he doesn't explicitly say that the ought here is a moral ought but I take it that given the context it is obviously meant to be a moral ought.

59. Joyce, *Myth*, 137.
60. Joyce, *Myth*, 177.
61. Notice how Railton only requires full *instrumental* rationality as opposed to Smith's full rationality where the demands of rationality can actually be quite substantive and will only emerge at the end of normative ethics so to speak. I'm not sure if Joyce fully appreciates this feature of Smith's view. In any case he never expresses any worries about instrumental rationality and Railton doesn't require, or even expect, convergence in what is non-morally good for us.
62. Railton, "Moral Realism," 162 n. 30.
63. For Railton's own take on this matter, see Railton, "Moral Realism," 140–41.
64. Joyce, *Myth*, 215.
65. Joyce, *Myth*, 214.
66. See Stanley, "Hermeneutic Fictionalism," for a discussion of some general problems with hermeneutical fictionalism. See Stephen Yablo, "Go Figure: A Path through Fictionalism," *Midwest Studies in Philosophy* 25 (2001) for some responses.
67. Stephen Yablo, "Abstract Objects: A Case Study," *Philosophical Issues* 12 (2002): 87.
68. Yablo draws on the work of Kendall Walton. See in particular Kendall Walton, *Mimesis as Make-Believe: On the Foundations of the Representational Arts* (Cambridge: Harvard University Press, 1990) and Kendall L. Walton, "Metaphor and Prop Oriented Make-Believe," *The European Journal of Philosophy* 1, no. 1 (1993).
69. I will not deal here with Donald Davidson's famous denial of metaphorical content (Donald Davidson, "What Metaphors Mean," in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984; originally published *Critical Inquiry*, 5 (1978), 31–47)). See David Hills, "Aptness and Truth in Verbal Metaphor," *Philosophical Topics* 25, no. 1 (1997) for a recent discussion and response.
70. I will be very selective, for reasons of space, about which features I will consider here. Consider Yablo's list of "suggestive similarities" between platonic objects and "creature of metaphorical make-believe" ("Paradox of Existence," 301–304). Moral properties will not share many of these suggestive similarities and it will take careful and controversial work to sort out which similarities they do share.
71. Stanley Cavell, "Aesthetic Problems of Modern Philosophy," in *Must We Mean What We Say?* (New York: Charles Scribner's Sons, 1969), 79. Quoted in Hills, "Verbal Metaphor," 122. As Hills points out, Cavell takes himself as following William Empson.
72. I am playing with Cavell's paraphrase of course (Cavell, "Must We Mean," 78–79).
73. It burns out the eyes of the beholder. Indeed the glance is so lethal that even the beloved cannot look at herself or himself in the mirror (D. J. Matthews and C. Shackle, *An Anthology of Classical Urdu Love Lyrics* (London,: Oxford University Press, 1972), 124–25). The gender is as usual ambiguous. In that and other ways the emotions of the traditional lover of the *ghazal* towards his beloved are of course quite different from those of Romeo.

    Another example from a different, and difficult, Ghalib verse: "It is from the ray of the sun that the dew learns the lesson of annihilation/I too exist only

till a glance of favour"—the glance of favour of the beloved no doubt (Mohammed Sadiq, *A History of Urdu Literature*, 2nd, rev. and enl. ed. (Delhi and New York: Oxford University Press, 1984), 274; translation modified).

74. Walton, *Mimesis as Make-Believe: On the Foundations of the Representational Arts*, 38.
75. Yablo, "Paradox of Existence," 290.
76. Yablo, "Paradox of Existence," 293.
77. Yablo, "Paradox of Existence," 295.
78. Yablo, "Paradox of Existence," 297.
79. See, for example, Stephen Yablo, "Does Ontology Rest on a Mistake?," *Aristotelian Society* Supplementary Volume 72 (1998): 250–55.
80. Yablo, "Does Ontology Rest on a Mistake?," 249.
81. Hills, "Verbal Metaphor," 147.
82. Hills, "Verbal Metaphor," 147.
83. The expression is used, and explained, by both Francis Snare and Stephen Darwall (see Francis Snare, *The Nature of Moral Thinking* (London and New York: Routledge, 1992), 6 and Stephen Darwall, *Philosophical Ethics*, ed. Norman Daniels and Keith Lehrer, *Dimensions of Philosophy Series* (Boulder: Westview Press, 1998), 8. See Jonathan Dancy, "Discussion on the Importance of Making Things Right," *Ratio* 17 (2004) for a recent deployment and explication.
84. Darwall, *Philosophical Ethics*, 8.
85. Yablo, "Does Ontology Rest on a Mistake?," 252. See also Richard Moran, "Seeing and Believing: Metaphor, Image and Force," *Critical Inquiry* 12 (16?) (1989).
86. Lewis, *On the Plurality of Worlds*, 2.
87. Earlier versions of this paper were presented to a conference on "Moral Theory after Nietzsche" at the University of Texas, Austin; to the research workshop "Social Ethics and Normative Theory" sponsored by the Mellon Foundation and the Stanford Humanities Center, Stanford University; and to the Department of Philosophy, Arizona State University. I thank the members of those audiences for very helpful comments. I would like to thank in particular Michael Bratman, Stewart Cohen, John Devlin, David Hills, Agnieszka Jaworska, Bernard Kobes, Brian Leiter, Robert Mabrito, Yonatan Shemmer, and Allen Wood. Ariela Tubert, the commentator for my presentation at the University of Texas, gave me very insightful and penetrating written comments. My thanks to Richard Joyce, Daniel Nolan, Greg Restall, Caroline West and Simon Blackburn for sharing their unpublished work on moral fictionalism with me.

References

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

Brandt, Richard B. *A Theory of the Good and the Right*. Oxford: Clarendon Press, 1979.

Bratman, Michael E. *Intention, Plans, and Practical Reason*. Cambridge: Harvard University Press, 1987.

———. *Faces of Intention: Selected Essays on Intention and Agency*. *Cambridge Studies in Philosophy*. Cambridge: Cambridge University Press, 1999.

Burgess, John P. "Why I Am Not a Nominalist." *Notre Dame Journal of Formal Logic* 24 (1983): 93–105.

Burgess, John P., and Gideon A. Rosen. *A Subject with No Object: Strategies for Nominalistic Interpretation of Mathematics*. Oxford and New York: Clarendon Press; Oxford University Press, 1997.

Cavell, Stanley. "Aesthetic Problems of Modern Philosophy." In *Must We Mean What We Say?*, 73–96. New York: Charles Scribner's Sons, 1969.

Crimmins, Mark. "Hesperus and Phosphorus: Sense, Pretense, and Reference." *Philosophical Review* 107, no. 1 (1998): 1–47.

Dancy, Jonathan. "Discussion on the Importance of Making Things Right." *Ratio* 17 (2004): 229–37.

Darwall, Stephen. *Philosophical Ethics. Dimensions of Philosophy Series*. Boulder: Westview Press, 1998.

Darwall, Stephen, Allan Gibbard, and Peter Railton. "Toward *Fin de siècle* Ethics: Some Trends." *The Philosophical Review* 101, no. 1 (1992): 115–89.

Davidson, Donald. "What Metaphors Mean." In *Inquiries into Truth and Interpretation*, 245–64. Oxford: Clarendon Press, 1984. Originally published in *Critical Inquiry*, 5 (1978), 31–47.

Evans, Gareth. *The Varieties of Reference*. Oxford: Clarendon Press, 1982.

Field, Hartry H. *Science without Numbers: a Defence of Nominalism*. Princeton: Princeton University Press, 1980.

———. *Realism, Mathematics, and Modality*. Oxford and New York: Blackwell, 1989.

Fine, Arthur. "Fictionalism." *Midwest Studies* 18 (1993): 1–18.

Garner, Richard. "On the Genuine Queerness of Moral Properties and Facts." *Australasian Journal of Philosophy* 68 (1990): 137–46.

———. "Are Convenient Fictions Harmful to Your Health?" *Philosophy East & West* 43, no. 1 (1993): 87–106.

Gibbard, Allan. *Thinking How to Live*. Cambridge: Harvard University Press, 2003.

Harman, Gilbert. "(Non-Solipsistic) Conceptual Role Semantics." In *Reasoning, Meaning, and Mind*. Oxford and New York: Clarendon Press; Oxford University Press, 1999.

Hills, David. "Aptness and Truth in Verbal Metaphor." *Philosophical Topics* 25, no. 1 (1997): 117–53.

Hudson, W. D. *Modern Moral Philosophy*. Garden City: Anchor Books, 1970.

Jackson, Frank. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford and New York: Clarendon Press, 1998.

Joyce, Richard. *The Myth of Morality*. New York: Cambridge University Press, 2001.

———. "Moral Fictionalism." *In Fictionalist Approaches to Metaphysics*, edited by Mark E. Kalderon. Oxford: Oxford University Press, forthcoming.

Korsgaard, Christine. "The Normativity of Instrumental Reason." In *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Gaut, 215–54. Oxford: Oxford University Press, 1997.

———. "Skepticism about Practical Reason." In *Moral Discourse and Practice: Some Philosophical Approaches*, edited by Stephen Darwall, Allan Gibbard and Peter Railton, 373–87. New York and Oxford: Oxford University Press, 1997.

Lewis, David K. *Counterfactuals*. Cambridge: Harvard University Press, 1973.

———. *On the Plurality of Worlds*. Oxford and New York: Blackwell, 1986.

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

Matthews, D. J., and C. Shackle. *An Anthology of Classical Urdu Love Lyrics*. London,: Oxford University Press, 1972.

Miller, Alexander. *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press, 2003.

Moran, Richard. "Seeing and Believing: Metaphor, Image and Force." *Critical Inquiry* 12 (16?) (1989): 87–112.

Nolan, Daniel, Greg Restall, and Caroline West. "Moral Fictionalism." MS, 2002.
Railton, Peter. "Naturalism and Prescriptivity." *Social Philosophy and Policy* 7, no. 1 (1990): 151–74.
———. "Moral Realism." In *Moral Discourse and Practice: Some Philosophical Approaches*, edited by Stephen Darwall, Allan Gibbard and Peter Railton, 137–63. New York and Oxford: Oxford University Press, 1997. Originally published in *The Philosophical Review* 95 (1986): 163–207.
Rosen, Gideon. "Modal Fictionalism." *Mind* 99 (1990): 327–54.
———. "What is Constructive Empiricism?" *Philosophical Studies* 74, no. 2 (1994): 143–78.
Sadiq, Mohammed. *A History of Urdu Literature*. 2nd, rev. and enl. ed. Delhi and New York: Oxford University Press, 1984.
Schroeter, Laura, and Francois Schroeter. "A Slim Semantics for Thin Moral Terms?" *Australasian Journal of Philosophy* 81, no. 2 (2003): 191–207.
Smith, Michael. *The Moral Problem*. *Philosophical Theory*. Oxford: Blackwell, 1994.
———. "Internalism's Wheel." *Ratio* 8, no. 3 (1995): 277–302.
Snare, Francis. *The Nature of Moral Thinking*. London and New York: Routledge, 1992.
Sobel, David. "Do the Desires of Rational Agents Converge?" *Analysis* 59, no. 3 (1999): 137–47.
Stanley, Jason. "Hermeneutic Fictionalism." *Midwest Studies in Philosophy* 25 (2001): 36–71.
Stevenson, C. L. "The Emotive Meaning of Ethical Terms." *Mind* 46 (1937): 14–31.
Van Fraassen, Bas C. *The Scientific Image*. Oxford: Clarendon Press, 1980.
Wallace, R. Jay. Review of Joyce, Richard, The Myth of Morality, Cambridge, 2002. *Notre Dame Philosophical Reviews* (2003).
Walton, Kendall. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge: Harvard University Press, 1990.
———. "Existence as Metaphor?" In *Empty Names, Fiction, and the Puzzles of Non-Existence*, edited by Anthony J. Everett and Thomas Hofweber, 69–94. Stanford, Calif.: CSLI Publications, 2000.
Walton, Kendall L. "Metaphor and Prop Oriented Make-Believe." *The European Journal of Philosophy* 1, no. 1 (1993): 39–56.
Wedgwood, Ralph. "Conceptual-Role Semantics for Moral Terms." *Philosophical Review* 110, no. 1 (2001).
Williams, Bernard. "Internal and External Reasons." In *Moral Luck: Philosophical Papers 1973–1980*, 101–13. Cambridge: Cambridge University Press, 1981.
———. "Internal Reasons and the Obscurity of Blame." *Logos* 10 (1989): 1–11.
———. "Replies." In *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, edited by J.E.J. Altham and Ross Harrison, viii, 229 p. Cambridge and New York: Cambridge University Press, 1995.
Yablo, Stephen. "Does Ontology Rest on a Mistake?" *Aristotelian Society* Supplementary Volume 72 (1998): 229–61.
———. "A Paradox of Existence." In *Empty Names, Fiction, and the Puzzles of Non-Existence*, edited by Anthony J. Everett and Thomas Hofweber, 275–312. Stanford, Calif.: CSLI Publications, 2000.
———. "Go Figure: A Path through Fictionalism." *Midwest Studies in Philosophy* 25 (2001): 72–102.
———. "Abstract Objects: A Case Study." *Philosophical Issues* 12 (2002): 220–40.