Fictionalism: Morality and metaphor
Richard Joyce

1. Introduction

Language and reflection often pull against each other. Ordinary ways of talking appear to commit speakers to ontologies that may, upon reflection, be deemed problematic for a variety of empirical, metaphysical, and/or epistemological reasons. The use of moral discourse, for example, appears to commit speakers to the existence of obligation, evil, desert, praiseworthiness (and so on), while metaethical reflection raises a host of doubts about how such properties could exist in the world and how we could have access to them if they did.

One extreme solution to the tension is to give up reflection—to become like those "honest gentlemen" of England, as Hume described them, who "being always employ'd in domestic affairs, or amusing themselves in common recreations, have carried their thoughts very little beyond those objects" (*Treatise* 1.4.7.14). Hume sounds a touch envious of the unreflective idyll, but the very fact that he is penning a treatise on human nature shows that this is not a solution available to him; and, likewise, the very fact that you are reading this chapter shows that it's unlikely to be a solution for you.

Another extreme solution to the tension is to give up language—not *in toto*, of course, but selectively: to stop using those elements of discourse that are deemed to have unacceptable commitments. How radical this eliminativism is will depend on the nature of the problem. We all stopped talking about mermaids when we discovered that there aren't any, and nothing very terrible ensued. (Of course, there isn't a moratorium on using the word "mermaid," but we have ceased to use the word in speech that commits us to their existence.) But some potentially erroneous ideas permeate our conceptual scheme far more extensively than thoughts of mermaids ever did. If, for example, we decided that there is something flawed in the idea of causation, or realized that folk psychology is a false theory, or came to the conclusion that moral thought is riddled with error, then the response of simply excising and jettisoning these areas of discourse (and all that imply them) seems hardly possible.

Most solutions peddled in the intellectual marketplace, however, shun these extremes, and rather pursue a compatibilist strategy of arguing either (A) that the language in question does not ontologically commit speakers in the way that it might appear to, or (B) that the commitments in question are not problematic after all—that is, that the world does contain the entities in question. The distinction between these two strategies is not always clear-cut. Regarding moral discourse, for example, the complaint has been voiced that in making moral judgments we commit ourselves to the existence of objective values, but that the world contains nothing so weird

(Mackie 1977). One response is to note that the concept of *objectivity* is ambiguous and indeterminate, and it can then be argued (following strategy A) that moral discourse was never committed to one form of objectivity—a kind of moral objectivity that the world does not satisfy—but (following strategy B) that moral discourse nevertheless is committed to another form of objectivity—a kind of moral objectivity that the world does satisfy. On this view there is no tension between moral language and reflection; rather, the appearance of tension is one of philosophers' own invention, of our being clumsy or too coarse-grained in identifying the conceptual commitments of moral discourse.

This chapter is focused on a kind of non-extreme solution to the tension—fictionalism—and will concentrate on the example that I have already been using: morality.

2. Moral fictionalism

For anyone feeling the tension between the apparent ontological commitments of moral discourse, on the one hand, and what reflection suggests there is in the world, on the other, the language of fiction should be of interest because it reveals various ways by which ontological commitment is nullified. There is no single means by which fiction accomplishes this. One way is the act of initiating a fiction using the cohortative mood: someone who says "Let's pretend that the sofa is a dragon" is not committed to the existence of dragons. A second way is when the fiction has been entered into and the utterance of a sentence is a move within that fiction: someone who, in the course of telling a story, says "And in this cave there lived a dragon" is not committed to the existence of dragons. A third way is when talking about a fiction: someone who says "In *The Hobbit* there's a dragon" is not committed to the existence of dragons. Another way is the familiar mini-fiction of metaphors: someone who says "My high school math teacher was a dragon" is not committed to the existence of dragons.

This list is not exhaustive, and different moral fictionalists have different proclivities. What unites them is that they seek to model our potentially problematic moral discourse on one or other commitment-nullifying aspect of fiction. But the extent to which one interprets a discourse as "similar to fiction" is unavoidably vague. For example, one way of interpreting moral discourse in a commitment-nullifying manner is old-fashioned emotivism, according to which "Stealing is morally wrong" (say) is taken to mean something like "Boo to stealing!" whereby all reference to the troublesome property of wrongness evaporates (Ayer 1936). Does such a view see moral discourse as similar to fictive discourse in any interesting way? No, I wouldn't think so, and so emotivism is not a form of fictionalism (though its motivation may be similar). By comparison, consider (just for illustrative purposes) the improbable view that interprets moral discourse as akin to telling jokes—joking being a familiar commitment-nullifying practice. Would such a view see moral discourse as similar to fictive discourse in any interesting way? It's hard to say. Jokes involving narratives and satire probably do count as a kind of fiction, whereas puns and witty insults

probably do not. So this "komoidist" metaethical theory (which I've just invented and proudly named) may or may not count as a form of fictionalism; it's borderline. So not only are there many varieties of moral fictionalism, but the boundaries of the family will be vague. However, I don't think that this is a particular problem or that there's any need to work to eradicate the vagueness. The variety of fictionalisms, by contrast, is something that we must continue to bear in mind, since what is an objection to one form of theory may not be to another, and what is a point in favor of one theory may not be for another.

Some fictionalists, for example, will see engaging in moral discourse as a kind of make-believe, but not all do. If moral discourse is modeled in the third way just listed, then make-believe has little to do with it, for someone who says "In *The Hobbit* there's a dragon" is not engaged in an act of pretense at all. (At most, she is making an assertion about a fictional work which might be engaged with via an act of make-believe.) Some fictionalists deny that moral judgments are beliefs and others do not. The sentence "Let's pretend that the sofa is a dragon" does not express a belief (i.e., is not an assertion), while the sentence "In *The Hobbit* there's a dragon" typically would be used to express a belief. Thus some moral fictionalists are noncognitivists and others are cognitivists. Those who are cognitivists leave the door open to the possibility of moral knowledge, while those who are noncognitivists do not; thus some moral fictionalists will count as moral skeptics and others will not.

Perhaps the most basic division among these many forms of fictionalism, however, is that between revolutionary and hermeneutic moral fictionalism. The revolutionary fictionalist doesn't think that our actual moral discourse is particularly like a fiction at all. Our actual moral discourse (she thinks) ontologically commits us to false things: the existence of objective values, perhaps, or the existence of pure free will, or the existence of truth-guaranteeing moral intuitions, or some such improbable or even impossible thing(s). The revolutionary fictionalist starts out life as a moral error theorist. But the revolutionary fictionalist doesn't like the look of the second extreme response mentioned above: of eliminating moral discourse altogether. Perhaps she perceives practical costs down that eliminativist route.[1] And so she recommends that we maintain our moral discourse but sans the ontological commitment, looking to some aspect of familiar fictive discourse as a model for how this nullification of commitment might be accomplished. The hermeneutic fictionalist, by contrast, argues that our moral discourse was never ontologically committed in the first place, and argues that the way moral discourse actually accomplishes this nullification is substantively similar to some means by which familiar fictive discourse accomplishes the same.

---

[1] In philosophy of mind, the term "eliminativism" has been used principally as a synonym of "error theory." The kind of "eliminativist materialism" advocated by Paul Churchland (1979, 1981) and Patricia Churchland (1986) is a claim about the non-existence of the kinds of folk psychology. But in metaethics the term "eliminativism" refers to a practical option that one might take (or not take) once one has already adopted the moral error theory. Roughly: the former banishes something from our ontology, the latter banishes something from our language. (See Joyce 2013 for some comparison.)

What the revolutionary fictionalist recommends we become, in other words, the hermeneutic fictionalist argues we already are. This fundamental difference makes it difficult to discuss them in one breath. One theory is an alternative to the moral error theory, while the other is a move that may be made by someone who has embraced the error theory. One theory must be assessed as true or false, while the other must be assessed as good advice or poor advice. For one we must consider evidence concerning how moral language is actually used, for the other we must contemplate counterfactuals concerning what costs and benefits might accrue from various linguistic and psychological transformations.

I noted earlier the inherent vagueness of fictionalism, inasmuch as it claims that moral discourse should be interpreted as "similar to fiction." There's another vagueness that should be mentioned: between the revolutionary fictionalist and the eliminativist opposition. The eliminativist doesn't recommend a total ban on moral vocabulary, any more than we ban the use of words like "mermaid," "dragon," or "phlogiston"; what the eliminativist bans is the use of such words in speech that commits us ontologically to the items in question. However, now we can see that, broadly speaking, the fictionalist is making the same recommendation: that we carry on using moral language but in an ontologically innocent manner. The fictionalist cannot even claim to be introducing some new way of nullifying commitment, for the position explicitly models itself on extant ways. And the eliminativist has no objection to moral language being used in fictive contexts: there is no ban on an actor reciting the lines "Not in the legions of horrid hell can come a devil more damn'd in evils to top Macbeth." Where the revolutionary fictionalist and the eliminativist diverge is concerning all those pre-revolutionary moral judgments that (they agree) commit speakers to a problematic ontology. The eliminativist thinks that these utterances must go; the fictionalist thinks that many of them can stay *sans commitment*. But the fictionalist needn't claim that we can keep *all* the pre-revolutionary moral discourse. Perhaps some of it cannot be usefully "fictionalized." So ultimately the eliminativist allows that *some* moral language can remain (in obvious fictions, and so forth), and the fictionalist argues that *much more* moral language can remain (in less obvious fictions, perhaps), but there's no principled or exact line to be drawn between the two.

All forms of moral fictionalism face problems, and I am not confident that all can be overcome. However, it seems to me that some of the objections to the view(s) are based on misunderstandings of what the moral fictionalist is claiming (or, at least, what he or she *could* claim), and it is the goal of this chapter to bring a couple of such misunderstandings into the light. One is an objection to hermeneutic fictionalism and one is an objection to revolutionary fictionalism. Since these two versions of fictionalism are contraries, it should be apparent that in defending both here it is not my ultimate position to advocate both. For the record, my sympathies lie more with the moral error theory than with hermeneutic fictionalism, and as an error theorist I am sensibly ambivalent between eliminativism and revolutionary fictionalism, since I think the counterfactuals too complex to pronounce on with confidence.

The criticism of hermeneutic fictionalism to be discussed is called "the phenomenological objection"; the criticism of revolutionary fictionalism to be discussed I shall call "the unconstrained fictions objection." I discuss these side by side in this chapter because I think that both objections can plausibly be punctured by reflecting on the same phenomenon: metaphor. Too often, it seems to me, opponents of fictionalism have reached for the wrong paradigm to frame the view. How could moral discourse be like playing Cowboys and Indians?! (asks Matti Eklund, somewhat tongue-in-cheek (2015)). But assuming that the fictionalist likens moral discourse to childish games of pretense is an uncharitable distortion, and it is no wonder that the theory immediately looks far-fetched. (That said, references to childish games can be very useful in illuminating aspects of pretense—as I am going to do myself very shortly!) Thinking, by contrast, of moral discourse along the lines of metaphorical language—as suggested by Steve Yablo's many discussions of fictionalism (2000, 2001, 2005)—delivers more plausible versions of fictionalism. This is not to say that moral discourse *is* metaphorical (nor that it is being recommended that it *should* be); rather, the claim is that reflection on how metaphor works helps dispel certain challenges.

3. Metaphor: some distinctions

Yablo follows Kendall Walton in arguing that metaphors are prop-oriented acts of make-believe. Prop-oriented pretense contrasts with content-oriented pretense. In the latter, it is really the fiction that is of interest, and the props act in its service. Children may want to play at being knights, and some sticks from the garden aid them in their game by serving as swords. But when their parents say "Don't bring your swords into the castle," they are exploiting the game in order to say something primarily about the props themselves: *No sticks in the house.* This is a tempting model for metaphor, for when using a metaphor, one typically says something false in order to draw attention to something true. When I tell you that my math teacher was a dragon, I am letting you know that she was bad-tempered and warranted nervousness among her students. The latter is the important information; the manifest falsehood is an evocative means I use to convey it. (To see that metaphors are only *typically* false, imagine my math teacher, upon hearing of my low opinion of her, protesting "I was *not* a dragon!" This is no less a metaphor than my original non-negated claim, but of course its manifest content is entirely true.)

Sometimes it is very clear what literal information stands behind a metaphor, perhaps especially when a metaphor has gathered a sufficiently large fan-base that it has become an idiomatic expression. "It's raining cats and dogs" simply means that it's raining very hard. But often it's difficult to say with precision what a metaphor is driving at, even though we're comfortable that we understand it. I had to pause and think about how to articulate what stands behind "My math teacher was a dragon" (and even so I'm not confident that I got it quite right). The hackneyed example of "Juliet is the sun" is even trickier to render literally (without recourse to further

metaphors, concerning her "radiance," etc,). And metaphors can get more complex still:

> Insidious cruelty is this
> that will allow the heart
> a scent of wild water
> in the arid land —
> that holds out the cup
> but to withdraw the hand […]
> ("Wild water" by May Swenson)

Although we may have a good "ball-park" understanding of what's being conveyed, one might think that aiming to articulate anything more precise than that would be an error. [2] There is truth behind the metaphor, but it's intentionally indeterminate. Incidentally, Swenson's poem includes another nice example of why metaphors aren't always false, for it would indeed be cruel to offer a cup of water to a thirsty person and then withdraw it; it's just that this isn't what Swenson's really talking about. But wondering what she is talking about *precisely*—a withdrawn offer of love? of sex? of hope?—is not the right thing to wonder.

Suppose my math teacher was bad-tempered and I say "She was a dragon." Suppose, by contrast, that my English teacher was amicable and I say "He was a dragon." The manifest contents of both claims are equally false, yet there is evidently something wrong with the second metaphor that is not wrong with the former. One might say that the second is both literally false *and also metaphorically false*, whereas the first is literally false but metaphorically true. Or one might prefer to say that they're both simply false, but the first is *apt* in the way the second is not.

Walton and Yablo argue that metaphor involves pretense (see also Hills 1997 and Egan 2008). Consider Romeo's metaphorical adoration of Juliet: it seems reasonable to say that Romeo pretends to assert one thing, but in fact says (or at least conveys) something else; and any competent listener understands that Romeo is engaging in or alluding to this make-believe (see Hills 1997). Catherine Wearing (2012) rejects this pretense view (see also Camp 2009). She thinks, for a start, that the view will have trouble accommodating *true* metaphors, for in these cases by what cue will the hearer know that an act of make-believe is underway? The usual cue would be the manifest falsity of the metaphor: Juliet is very clearly *not* the sun, and Romeo very clearly knows this (or else his foremost concern would surely be his imminent vaporization). In the case of my math teacher objecting "I was *not* a dragon," the cue is presumably that this sentence, although true, stands in no need of assertion—its truth is too obvious and not salient. Wearing worries about the case of someone looking at storm

---

[2] The literary critic Cleanth Brooks wrote: "Let the reader try to formulate a proposition that will say what the poem 'says.' As his proposition approaches adequacy, he will find, not only that it has increased greatly in length, but that it has begun to fill itself up with reservations and qualifications … The truth is that all such formulations lead away from the center of a poem—not toward it" (1947: 198-99).

clouds on the horizon, saying "There's a storm coming," but alluding to some foreseen unpleasant events (of a non-meteorological nature). The hearer (she thinks) will receive neither a cue of falsity nor a cue of inappropriateness in order for the metaphorical interpretation to be triggered. It seems to me, however, that this is exactly the kind of scenario where the metaphor might be expected to misfire. If commenting on the pending storm is, in the conversational context, a natural thing to say, then the intended metaphorical allusion may be lost on the hearer. If, on the other hand, the focus of preceding conversation had been the foreseen unpleasant events, and if a sudden comment about the weather would be (though true) unexpected, then the hearer may catch on to the metaphor. We can easily imagine the scenario such that the hearer isn't sure whether the cue's been given (perhaps the speaker didn't get his/her tone of voice quite right), in which case the speaker may feel confused about whether any game of make-believe is afoot. This doesn't seem a problem for the pretense view; sometimes metaphors backfire.

Wearing goes on to worry about what the hearer is supposed to be *pretending* in this example. How can he pretend that there's a storm coming when there *is* a storm coming (and he knows it)? She writes: "It's a bit like imagining one's cat to be sitting in the window while watching one's cat sitting in the window" (2012: 513). But this seems to misidentify the act of pretense. Remember that what Romeo pretends is that *he is asserting that* Juliet is the sun (see Walton 2014). Similarly, one who says "There is a storm coming"—on a sunny day and in reference to foreseen unpleasant non-meteorological events—is pretending to assert that there is a storm coming. But this remains true on a non-sunny day, when there really is a storm coming. Of course, the speaker may simply have asserted the sentence in a non-metaphorical way. But if the speaker has infused the utterance with metaphorical allusion, then he is *not* simply asserting the sentence; he is pretending to do so.[3] When an actor playing Hamlet recites the line "It is very cold," he is pretending to assert this; and this remains no less true if the play happens to be performed on a chilly evening.

I have made some brief comments in defense of the so-called pretense view of metaphor, but ultimately my arguments in this chapter don't depend on its success. I have already noted that certain versions of fictionalism do not privilege the idea of make-believe, and it is moral fictionalism that I am interested in supporting, not a specifically make-believe-oriented version of moral fictionalism.[4] I shall, however, continue to talk as if make-believe is a central concern—for the reason that it *is* the central concern of prominent versions of fictionalism and the focus of much of the criticism that I hope to deflect.

4. The phenomenological objection to moral fictionalism

---

[3] It's a tricky case because the speaker is trying to do something "clever": to use a single utterance to play between two different speech acts: a literal one concerning the weather and a metaphorical one concerning pending unpleasantness. The savvy hearer will *know* that the utterance is being used to do something unorthodox and clever, and may take pleasure in the ambiguity.

[4] Toward the end of her paper, Wearing acknowledges versions of fictionalism that don't involve make-believe (2012: 519-20).

Several critics of hermeneutic fictionalism (not necessarily of the moral variety) have found it implausible on the ground that it simply doesn't *seem* as if we're dealing with a fiction when we use the discourse in question. (See Stanley 2001: 46; Eklund 2007; Chrisman 2008: 7; Reynolds 2009: 315.) David Liggins sums it up nicely:

> One commonly voiced objection [to hermeneutic fictionalism] is that we do not seem to be pretending when we engage in the discourses of which pretense theories are claimed to be true: indeed we seem not to be pretending. For instance, if we introspect, it seems to us that we believe, rather than pretend, that $2 + 2 = 4$. (Liggins 2010: 768)

When the paradigm of engaging with a fiction is something like playing cops and robbers with the kids, or watching a play in a theater, then the phenomenological objection may have some bite. After all, no matter how emotionally immersed in the play we are, we don't lose sight completely of the fact that it's all fake. (When the hero is killed, we keep munching popcorn rather than rush to phone the police.) One would have to conjure up a fairly uncommon scenario in order to describe a person who is genuinely unsure whether s/he is witnessing a play or experiencing reality.

But when the model of engaging with fiction is the use of metaphor, the phenomenological objection loses most if not all of its bite. We constantly use metaphors without being aware that we're doing so, and there are many instances where most of us wouldn't be quite sure, even after reflection, whether a phrase counts as a metaphor or not. Metaphors often become frozen in language, and perhaps at some point they cease to be metaphors. But it is very hard to say at what point, and it is hard to know quite what happens to a word or phrase, beyond a kind of entrenched familiarity, that might cause it to lose its metaphorical quality.

Donald Davidson (1978) proposes that a metaphor dies when it no longer prompts the hearer to *notice* a similarity relation. (See also Orwell 1946.) He uses the example of the phrase "the mouth of a river." Perhaps there was a period in the mists of history when the phrase conjured in the hearer's mind a fleeting image of a human's or animal's mouth, but that time has long past, and we now treat the word "mouth" as simply denoting (inter alia) the egress of a river. Davidson contrasts this with Homer's image of battle wounds as "mouths"—a striking expression that still prompts visceral comparative imagery.[5]

Walton (1993: 48) thinks that many so-called dead metaphors are not dead in the sense that Davidson means: that they've ceased to be metaphors. One of Walton's examples of a dead metaphor is using "male" and "female" for electrical and plumbing connections. There is still a fiction here, he thinks, though hardly one that

---

[5] I confess that I cannot actually find this metaphor in Homer (who greatly prefers similes to metaphors), but this hardly matters. There is a gruesome example of the simile version from Michael Drayton: "So that their wounds, like mouths, by gaping wide, / Made as they meant to call for present Death, / Had they but Tongues, their deepness gives them breath" (*The Barons' Wars* (1603)).

one is likely to be emotionally caught up in—after all, in this prop-oriented make-believe we're not much interested in the game itself. While "the conscientious plumber does his job without, fictionally, leering at the fixtures" (Walton 1993: 40), the very fact that the plumber might feel embarrassment at explaining the terminology to a child shows that there remains some life in the metaphor. And there does seem to be something overly demanding in Davidson's criterion, for surely a metaphor doesn't have to *actually* prompt a comparative image in the hearer; it should be enough that it *can*. But then all those dead metaphors rise from their graves, for even something as dull and familiar as "the mouth of the river" *can*, with a small effort, be used to conjure a comparative image—in a way that, say, "the salinity of the river" cannot.

It is not my intention to settle or even take sides in this debate here. The fact that debate exists is sufficient to make my point. Two intelligent and educated people could disagree about whether "the mouth of the river" remains a metaphor (or if not this example, then a myriad of others), and would thus (by certain lights) disagree about whether a user of the phrase is involved in make-believe. Their disagreement would extend even to their own uses of the phrase: one could not reasonably claim to have first-person authority regarding the matter, discernable through introspection of one's phenomenology. Rather, the matter would have to be settled through careful and critical examination of competing theories of metaphor; it is a theory-laden truth. If this "careful examination" is something one has not yet undertaken, or doesn't care to, then one may remain indefinitely in a state of indecision regarding whether using a phrase involves one in make-believe. Thus, unlike regarding the person watching a play, one does not have to conjure up an uncommon scenario in order to describe a person who is genuinely unsure whether s/he is speaking figuratively or speaking literally; I submit that it's a commonplace state.

There's a slightly different way of making the same point. Earlier I sketched the debate between those who hold a "pretense theory" of metaphor (Walton, Yablo, Hills) and their opponents (Wearing, Camp). Consider something that's obviously a live metaphor: say, Churchill's description of Mussolini as "a utensil" in his 1941 address to the US Congress. The question now is not whether this is a metaphor—it most certainly is a very lively one—but whether it involves make-believe. Again, the fact that debate exists is sufficient to make my point. The matter won't be settled by introspection of one's phenomenology; it's a theory-laden truth.

In short, the phenomenological objection to fictionalism is dulled by noting that while some acts of make-believe have a fairly obvious phenomenological component, others do not. Or if we're to put the point in a way that side-steps reference to make-believe: While some acts of engaging with fiction have a fairly obvious phenomenological component, others do not.

The phenomenological objection is leveled at hermeneutic fictionalism rather than revolutionary fictionalism because the revolutionary fictionalist can always allow that taking a fictive attitude toward morality *would* have a different phenomenology than having an ontologically-committing doxastic attitude toward it. But there is, nevertheless, an objection to revolutionary fictionalism that is related to the

phenomenological objection: namely, the worry that if the post-revolutionary ontologically non-committing fictive moral discourse does *feel different* than the pre-revolutionary committing moral discourse—if it "feels like make-believe"—then how are we to take it seriously as the basis for important practical decisions? If a central part of the point of morality is to motivate us to act in ways that we might otherwise be tempted not to, then surely its having the phenomenology of pretense would undercut its ability to play such a role? The best way of responding to this worry will emerge in my discussion of the second objection to fictionalism.

5. The unconstrained fictions objection to moral fictionalism

The revolutionary fictionalist holds that our current discourse embroils speakers in systematic error, but the error can be avoided by adjusting the discourse (either in its content or our attitude toward it) so as to resemble a non-committing fictive engagement. Several critics of revolutionary fictionalism (not necessarily of the moral variety) have found it implausible on the ground that there is a multitude of possible fictions and no non-arbitrary way of choosing among them. (See Sainsbury 2009: 204; Chrisman 2017: 60; and Nolan et al. 2005: 327 for discussion.) Christopher Peacocke accuses modal fictionalists of *fetishism* if they cannot answer the charge:

> The charge of fetishism lodged against the modal fictionalist is that there is no saying what is so special about the theory PW of possible worlds which is mentioned in the schema asserted by the modal fictionalist. Since the fictionalist does not take the theory PW to be true, why should we be peculiarly interested in what follows from it, rather than from some other theory? (Peacocke 1999: 154)

When the paradigm of engaging with a fiction is something like playing cops and robbers with the kids, or watching a play in a theater, then the problem of unconstrained fictions may seem persuasive. After all, the action of such a game or the narrative of a play could take all sorts of directions. Even at the play's intermission, when we know half the storyline, there's an infinite number of ways the rest of the play could continue that are consistent with what we know so far.

But when the model of engaging with fiction is the use of metaphor—especially when modeled on prop-oriented make-believe—then the unconstrained fictions objection recedes. Prop-oriented pretense is constrained by the pertinent information one seeks about the prop. I like Walton's example of holding up a boot to represent Italy, in order to show someone the location of the town of Crotone (1993: 40). "Crotone is *here*," I say, pointing to a spot on the sole. This is a false claim; I do not have a town of over 50,000 inhabitants on my boot. Yet I pretend to assert that I do in order to initiate or refer to a kind of make-believe. But it is not the make-believe per se with which my hearer and I are concerned; we're concerned with some real geographical information, and my false statement (along with my pointing) is heavily constrained by this end. Had I instead pointed to the ankle of my boot when indicating

the location of Crotone, the whole exchange would have backfired and my audience would rightly feel misled.

The moral fictionalist is not recommending that we make a fiction of morality for any old airy-fairy reason, any more than using a metaphor amounts to uttering a random falsehood because it's gratifying to do so. We make a fiction of morality because it serves our purposes, and thus it matters enormously *which* moral fiction (of the infinite number of possibilities) we adopt. To choose the wrong fiction could be ill-advised or disastrous.

If moral discourse were to become a constrained fiction, then, by what facts would it be constrained?

A preliminary to addressing this question is to remember to whom the fictionalist advice is being offered (i.e., to whom does the "our" refer in "it serves our purposes"?).[6] Many of the pieces of fictionalist advice amount to altering linguistic conventions, and therefore cannot be sensibly offered to individuals, any more than one might offer an individual the advice "Become an orchestra." We must see the revolutionary fictionalist as offering advice to a *group*. This raises the question of what sense can be made of certain courses of action serving the purposes of a group.

Bearing in mind that revolutionary fictionalists are moral error theorists, the answer that will likely appeal to them (though it is by no means mandatory) is a broadly Humean one. An individual's purposes are a function of his/her desires (perhaps with a dose of idealization), and the purposes of the group will be some function of its members' purposes. (It is not lost on me that this leaves a great deal unspecified; more on this in a moment.) Let us assume that the moral error theorist is not an error theorist about such Humean reasons.

It follows that the error theorist insists that there is something more to moral normativity than Humean resources can supply. I don't have space on this occasion to go into what arguments might undergird this insistence, but their general shape is well-known. (See Joyce 2001, 2011a, 2011b; Olson 2014.) A conspicuous premise in such arguments will likely be the thought that moral normativity is imbued, conceptually speaking, with more practical authority (what Mackie famously calls "to-be-pursuedness" and "not-to-be-doneness (1977: 40)) than can be provided by the Humean system. The whole point of morality (this thinking goes) is to prescribe and proscribe actions in a desire-independent manner.[7] Put simply: Kant was right about the nature of moral concepts, and Hume was right about the nature of both practical rationality and the world.[8]

---

[6] For discussion of this issue with respect to hermeneutic fictionalism, see Brock 2014.

[7] These arguments, if they are sound, would defeat the natural rejoinder that the revolutionary fictionalist is undermining her own error theoretic foundation in accepting what might be called "Humean normativity." If one is willing to be a success theorist about Humean normativity (the rejoinder runs) then why cannot one build moral facts out of these resources? (See, e.g., Lenman 2008.) It's a perfectly good question, but my point is that this is not the time to raise it. In assuming (as we are) that the revolutionary fictionalist takes herself to have grounds for being a moral error theorist in the first place, we are assuming that she takes herself to have grounds for rejecting this rejoinder.

[8] I am in particular using references to "Kant" merely as a figurehead, rather than making any serious attribution of a metaethical view to him.

One way, then, to interpret revolutionary fictionalism is as the recommendation that we maintain our (false) Kantian moral discourse as a way of conveying information about our (true) Humean reasons. The Humean reasons provide an anchor that constrains the moral fiction. Assume, say, that I have a Humean reason not to break a certain promise. Both the propositions "I am morally obliged to keep the promise" and "I am morally permitted to break the promise" are false.[9] However, the former falsehood can convey true information about the Humean reasons in a way that the latter cannot. As with the earlier example of using a prop incorrectly in prop-oriented make-believe, were I to say "I am morally permitted to break the promise," then the exchange would backfire and my audience would rightly feel misled. This is why moral fictions can still be taken seriously as the basis of important decision-making: because they function like metaphors for the serious matters that do form the basis of sound decision-making.

What would be the advantages of speaking falsely in this fashion? Wouldn't it be better (simpler, more honest, less misleading, less confusing) to talk literally in the Humean vocabulary—i.e., to embrace eliminativism? Again, reflection on metaphor is useful here. Metaphors can often create an impact when the literal talk will leave us cold—an impact on emotions and (thus) on motivations. Metaphors draw attention to certain aspects of the true target of discussion, masking off attention to other aspects. It is not implausible to suspect that Humean reasons are vulnerable to forms of practical self-sabotage (e.g., weakness of will) precisely because of their ultimate desire-dependency (see Joyce 2001: 136-7, ch. 7). An expedient that draws attention away from this feature by making reference to objective desire-independent reasons may better serve our practical purposes (our real Humean purposes, that is).

I have been talking as if there were a precise Humean normative framework which constrains the moral fiction—as if we could in principle do away with the fiction and speak in literal terms except that this would be somehow (contingently) practically self-undermining. My final point is to note that this may be an idealization. An important feature of metaphors is that we often *cannot* locate a determinate fact lying behind them. Metaphors often serve their purpose perfectly well—indeed, oftentimes, *better*—by resisting paraphrase. We may not be able to articulate precisely what facts are conveyed by "Juliet is the sun," but the metaphor still remains firmly anchored in real properties and relations that Juliet instantiates. Similarly, the Humean framework of group purposes is indeterminate along various dimensions. An individual's purposes (/ends /values /reasons) may be a function of his/her desires, but *which* function? There are a number of options here, and it is not clear that there's a fact of the matter about which is correct. And a group's purposes may be a function of the purposes of its members, but, again, *which* function? There are many options here, too. None of this, it seems to me, speaks against the possibility that employing a moral fiction may be a useful way of conveying information about the Humean world.

---

[9] They may not even make sense. Effective metaphors often contain category mistakes (e.g., "…hateful thoughts enwrap my soul in gloom" (from Keats' *To Hope*)).

6. Conclusion

There are many more challenges to moral fictionalism than I have touched on here; I do not claim even to have tackled the most serious objections. Nevertheless, I hope to have shown that some common and natural objections to moral fictionalism are usefully approached by reflecting on the familiar phenomenon of metaphor.

References

Ayer, A. J. 1936. *Language, Truth and Logic* (London: Victor Gollancz).

Brock, S. 2014. "The phenomenological objection to fictionalism." *Philosophy and Phenomenological Research* 88: 574-592.

Brooks, C. 1947. *The Well-Wrought Urn: Studies in the Structure of Poetry* (New York: Reynal & Hitchcock).

Camp, E. 2009. "Two varieties of literary imagination: Metaphor, fiction, and thought experiments." *Midwest Studies in Philosophy* 33: 107-130.

Chrisman, M. 2008. "A dilemma for moral fictionalism." *Philosophical Books* 49: 4-13.

Chrisman, M. 2017. *What is this Thing Called Metaethics?* (NY: Routledge).

Churchland, Patricia. 1986. *Neurophilosophy: Toward a Unified Science of the Mind/Brain* (Cambridge, MA: MIT Press).

Churchland, Paul. 1979. *Scientific Realism and the Plasticity of Mind.* (NY: Cambridge University Press).

Churchland, Paul. 1981. "Eliminative materialism and the propositional attitudes." *Journal of Philosophy* 78: 67-90.

Davidson, D. 1978. "What metaphors mean." *Critical Inquiry* 5: 31-47.

Egan, A. 2008. "Pretense for the complete idiom." *Noûs* 42: 381–409.

Eklund, M. 2015. "Fictionalism." In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. URL = < https://plato.stanford.edu/entries/fictionalism/>.

Hills, D. 1997. "Aptness and truth in verbal metaphor." *Philosophical Topics* 25: 117–153.

Joyce, R. 2001. *The Myth of Morality* (Cambridge: Cambridge University Press).

Joyce, R. 2011a. "The error in 'The error in the error theory.'" *Australasian Journal of Philosophy* 89: 519-534.

Joyce, R. 2011b. "The accidental error theorist." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Vol. 6* (Oxford University Press): 153-180.

Joyce, R. 2013. "Psychological fictionalism, and the threat of fictionalist suicide." *Monist* 96: 517-538.

Lenman, J. 2008. "Against moral fictionalism." *Philosophical Books* 49: 23-32.

Liggins, D. 2010. "The autism objection to pretence theories." *Philosophical Quarterly* 60: 764–782.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong* (NY: Penguin).

Nolan, D., Restall, G., & West, C. 2005. "Moral fictionalism versus the rest." *Australasian Journal of Philosophy* 83: 307-330.

Olson, J. 2014. *Moral Error Theory: History, Critique, Defence* (Oxford: Oxford University Press).

Orwell, G. 1946. "Politics and the English language." *Horizon* 13: 252-265.

Reynolds, S. 2009. "Making up the truth." *Pacific Philosophical Quarterly* 90: 315–335.

Sainsbury, R. M. 2009. *Fiction and Fictionalism* (NY: Routledge).

Stanley, J. 2001. "Hermeneutic fictionalism." *Midwest Studies in Philosophy* 25: 36–71.

Walton, K. 1993. "Metaphor and prop oriented make-believe." *European Journal of Philosophy* 1: 39–56.

Walton, K. 2000. "Existence as metaphor?" In A. Everett & T. Hofweber (eds.), *Empty Names, Fiction, and the Puzzles of Non-Existence* (Stanford, CA.: CSLI Publications) 69-94.

Walton, K. 2014. "Metaphor, fictionalism, make-believe: Response to Elisabeth Camp." Unpublished: URL = <http://global.oup.com/us/companion.websites/9780195098723/pdf/Metaphor_Fictionalism_Make_Believe.pdf>.

Wearing, C. 2012. "Metaphor, idiom, and pretense." *Noûs* 46: 499-524.

Yablo, S. 2000. "A paradox of existence." In A. Everett & T. Hofweber (eds.), *Empty Names, Fiction, and the Puzzles of Non-Existence* (Stanford, CA.: CSLI Publications) 275–312.

Yablo, S. 2001. "Go figure: A path through fictionalism." *Midwest Studies in Philosophy* 25: 72–102.

Yablo, S. 2005. "The myth of the seven." In M. Kalderon (ed.), *Fictionalism in Metaphysics* (Oxford: Oxford University Press) 88–115.