

## *Fictionalism in metaethics*

**Richard Joyce**

Penultimate draft of paper appearing in D. Plunkett & T. McPherson (eds.),  
*Routledge Handbook of Metaethics* (Routledge, 2017) 72-86.

### **1. Introduction**

Looking back, it is clear that humans have held massively mistaken beliefs about virtually every aspect of the world: the place of the Earth in the universe, the physical nature of everyday objects, where living things come from and how they reproduce, what happens to us when we die, and so forth. We are pretty good at getting things wrong. Often these errors can be corrected while the key concepts are revised rather than rejected. We didn't decide that the Earth, stars, animals, death, etc., don't exist; rather, we rectified our false beliefs about them and carried on talking about these topics (now, hopefully, more truly). Sometimes, however, the errors are so entrenched that the concepts in question seem beyond salvaging. Regarding angels, vitalistic life force, karmic reincarnation, tapu, supernatural divination, phlogiston, astrology, and so on, we didn't simply undertake an internal correction to the concepts in question, but rather decided that the whole conceptual framework in question was faulty.

It will be widely agreed that humans have held massively mistaken beliefs about *morality*. These mistakes may pertain to the substantive content of morality (believing that women should be subservient to men, for example) or to the general nature of morality (believing that God's will determines moral properties, for example). (One might try to reduce the extent of moral error by plumping for a form of radical relativism, but then one would have to ascribe false beliefs about morality to all those absolutists.) Many philosophers maintain (or hope) that these mistakes in moral thinking are of the former variety, such that our moral concepts can be patched up and we can carry on talking (now, hopefully, more truly) about right and wrong, virtue and vice, obligations and responsibilities, and so on. But the worry lurks in the background—and is sometimes embraced—that the mistakes are actually of the latter kind, and that moral thinking is a fundamentally flawed way of conceptualizing the world and ourselves. The moral error theorist maintains that moral facts belong on the list including angels, karmic reincarnation, phlogiston, and the rest. (Moral error theorists include Mackie 1977, Joyce 2001, and Olson 2014; see Olson, this volume.)

Fictionalism can be thought of as a way of trying to rescue morality from the threat of error theory. But fictionalism comes in different stripes, forcing the need to delineate upon anyone wishing to discuss it. The first distinction has come to have the somewhat unfortunate pair of labels “revolutionary” versus “hermeneutic.”

The revolutionary fictionalist thinks that the moral error theory is correct: our moral discourse really does involve systematic falsehood from which first-order moral truths cannot be salvaged. The usual view for the error theorist to take is to see actual moral discourse as ontologically committed to entities (e.g., moral properties) that do not exist. But of course not all language that involves non-denoting terms is

ontologically committing. The sentences “Angels do not exist” and “Aquinas believed in angels” are ontologically innocent with regard to angels. Certain utterances surrounding fiction also do not ontologically commit the speaker. Neither talking about a story—e.g., asserting “According to *It’s a Wonderful Life*, angels exist”—nor telling a story—e.g., uttering “Angels definitely exist!” without assertoric force in the course of a bedtime tale—commits the speaker to the existence of angels. The revolutionary fictionalist recommends that we carry on using moral discourse but in a manner that does not ontologically commit speakers to the problematic entities. The view is *revolutionary* in that it proposes an actual change in our attitude to morality: that we should alter our mental and linguistic lives so as to become ontologically innocent with respect to the problematic entities to which ordinary participation in moral discourse currently commits us. And the view is *fictionalist* if it turns to one or more of the familiar commitment-removing devices of ordinary fiction as a model for that change.

The hermeneutic fictionalist, by contrast, is not an error theorist about our actual moral discourse, but is nevertheless likely to be someone who is alert to the threat of error theory and therefore motivated to interpret our actual moral discourse in a manner that avoids that threat. Like the error theorist (and, indeed, like the noncognitivist), the hermeneutic fictionalist thinks that moral judgments are problematic if taken at face value. One who makes the judgment “Breaking promises is morally wrong,” for example, appears to assert that promise-breaking instantiates a certain property which (for a host of reasons familiar to metaethicists) proves to be metaphysically and epistemologically troublesome. So the hermeneutic fictionalist (like the noncognitivist) suggests that we therefore do not take moral judgments at face value—rather, an interpretation is offered such that discourse that appears to be problematically ontologically committed is not really thus committed. The view is *hermeneutic* in that it offers an interpretation of our actual moral discourse. And the view is *fictionalist* if it turns to one or more of the familiar commitment-removing devices of ordinary fiction as a model for that interpretation.

Note that I am restricting discussion to a fictionalist stance toward morality as a whole. One might, however, be a fictionalist about only certain parts of morality. For example, one might be a fictionalist about human rights but not about moral vices, or about evil but not about moral badness, and so on. Such possibilities of selective moral fictionalism are being put aside on this occasion.

What the revolutionary fictionalist recommends we become, the hermeneutic fictionalist declares we already are. Thus, though there is a great deal to be said about moral fictionalism that pertains to both revolutionary and hermeneutic forms, it is obvious that the two types of theory must ultimately be assessed in fundamentally different ways. Hermeneutic fictionalism purports to be *true*; we must therefore evaluate the evidence for and against it. Revolutionary fictionalism purports to be *good advice*; we must therefore evaluate its practical costs and benefits. Let us consider the prospects of the hermeneutic form first.

## Hermeneutic moral fictionalism

The hermeneutic fictionalist maintains that our actual moral discourse should be interpreted in a manner similar in some fashion to familiar fictional discourse. We have already seen two importantly different ways in which this “similarity” might be cashed out. On the one hand, there is what one does when one *talks about* a fiction; on the other hand, there is what one does when one *engages with* the fiction. The former view is type of cognitivism, according to which moral discourse is genuinely assertoric, it is just that the assertions concern the content of a fiction (e.g., “According to the fiction of morality, such-and-such is wrong”). This view therefore also allows for moral beliefs, moral truths, and (potentially) moral knowledge. The latter view is best construed as a type of noncognitivism, since what we do when we engage with a fiction (most obviously when telling a story) is not assert but *make-believe* that we assert. “Once upon a time...” is a device for showing that assertoric force is being lifted; “...and they lived happily ever after” is a device indicating its reinstatement; what comes between is generally amenable to a noncognitivist analysis. Note, though, that the noncognitivist analysis just described is one pertaining to the pragmatics of speech, not its semantics. A sentence within the story—e.g., “In this land, angels existed”—requires no translation into some special nondescriptive format; it means whatever it would mean if it were asserted; it is just that here it is uttered without that assertoric force. Some elements of story-telling, however, should be given a noncognitive analysis at the level of semantics rather than pragmatics. When one introduces an act of make-believe with the sentence “Let’s pretend that...,” the sentence is in the cohortative rather than indicative mood; it is not used to make an assertion. This raises the possibility of a hermeneutic fictionalism that interprets ordinary moral judgments as having the noncognitive logical form of “Let’s pretend that...” sentences (or some close cousin). (What I am calling “noncognitive fictionalism” might also be called “force fictionalism,” and what I am calling “cognitive fictionalism” might be called “content fictionalism. See Eklund 2015.)

The chief theoretic virtue enjoyed by cognitivist and noncognitivist versions of hermeneutic moral fictionalism alike is that they permit speakers to reap the benefits of moral discourse without footing the ontological bill for problematic entities. But is this advantage sufficient to motivate either view, and do either or both of the views suffer countervailing problems?

The cognitivist version of hermeneutic fictionalism faces some special problems. A sentence employing a story operator (“According to fiction *F*,...”) makes sense only to the extent that some account of the content of the story/fiction in question is forthcoming. (See Hussain 2004.) But even for paradigm fictions—such as the Sherlock Holmes stories, say—matters are far from straightforward. Consider the following:

1. According to the Holmes stories, Holmes lived on Baker Street.
2. According to the Holmes stories, Holmes had ten fingers.
3. According to the Holmes stories, Watson had forty-six chromosomes.

4. According to the Holmes stories, Watson had an old war wound in his shoulder.

1 is surely fine, and 2 seems safe despite the fact that the number of Holmes's fingers is never explicitly mentioned by Conan Doyle. 3, however, begins to feel problematic. And 4 is puzzling in a special way, since the position of Watson's war wound varies in different stories. However such matters should be straightened out, maintaining a degree of indeterminacy in the right places seems desirable.

But what "story" is the cognitivist hermeneutic moral fictionalist going to invoke in constructing the all-important operator? It's one thing to say, as the error theorist does, "Morality is just a fiction"; it's quite another to suppose that the fiction has sufficient determinate content to underwrite claims of the form "According to the fiction of morality..." More on this in a moment.

Another problem for the cognitivist hermeneutic moral fictionalist is accounting for how moral claims logically interact with non-moral claims, if the former but not the latter contains a tacit story operator. The following seems valid:

- P1: Stealing is morally wrong.  
P2: Amy stole last Tuesday.  
C: Therefore Amy did something morally wrong last Tuesday.

But if premise 1 harbors a tacit "According to the fiction of morality..." prefix, then the validity evaporates. Perhaps the validity could be rescued by adding the same prefix to both P2 and C, but the problem now is that the revised P2—"According to the fiction of morality, Amy stole last Tuesday"—seems simply false. (See Vision 1994.)

This problem can potentially be solved by understanding the relevant story operator more carefully. The "fiction of morality" differs from the Sherlock Holmes stories in that it is not a well-defined set of propositions. Rather, it is an image of the world—*this* world—as containing certain properties that in fact the world does not contain. It is a *fiction* not merely because it makes reference to non-actual entities, but because the entities may not even be *possible*. (See Proudfoot 2006.) Thus the story operator might be better rendered as "In fictional world *FW*,..."—accompanied with the reminder that fictional worlds are not possible worlds. (Exhibit A: Watson's war wound. Exhibit B: virtually any story involving time travel.) Despite the fact that we are not discussing possible worlds, it is reasonable to suppose that we can make some sense of roughly ordering fictional worlds with respect to their similarity to the actual world. The "moral *FW*" can be considered a complete world very much like the actual world but containing moral properties. (The moral fictionalist might borrow a device that Gideon Rosen employs in his discussion of modal fictionalism, according to which the *FW* contains an "encyclopedia" of non-moral truths. See Rosen 1990: 335.)

Understood in this manner, the aforementioned problem of how propositions with story operators logically interact seems more tractable. In particular, while "According to the fiction of morality, Amy stole last Tuesday" seemed false, a differently-worded revision of P2—"In the moral *FW*, Amy stole last Tuesday"—

stands a very good chance of turning out true. If in the actual world Amy did in fact steal last Tuesday, then, *ceteris paribus*, in the moral *FW* Amy stole last Tuesday.

Of course, what this solution requires is that all a speaker's utterances—even those that have nothing to do with morality—be interpreted as tacitly prefixed with “In the moral *FW*...” (since at any time a speaker might consider the logical relations between her non-moral claims and her moral claims). Whatever problems there may be with this prospect, there is nothing ridiculous *in general* with the idea that a great many of our utterances should be interpreted as bearing tacit prefixes. Consider how we would usually accept that most assertions make tacit reference to how things stand in the actual world rather than some other possible world.

But the first problem remains: What is the content of the moral fictional world? We can offer some general answers like “In the moral *FW*, moral obligations exist” and “In the moral *FW*, people have moral rights,” but can we hope for anything of more substance, like “In the moral *FW*, stealing is generally morally wrong”? When two people engage in moral argument—one of them claiming (we'll assume) that in the moral *FW* euthanasia is permissible, and the other claiming that in the moral *FW* euthanasia is not permissible—then we face not merely an epistemological problem of how to know which party is correct, but a far more serious problem (for this kind of fictionalist) of puzzlement over what it would even take for one party to be correct and the other incorrect.

What the cognitivist hermeneutic fictionalist evidently needs is some non-arbitrary means of restricting possible moral fictional worlds, ideally reducing the infinitude of candidates down to a single privileged fiction. (For arguments that the fictionalist need to pare fictions down merely to a *range* rather than a single fiction, see Woodward 2011.) Reflecting this desideratum, the story operator might be rendered “In the best moral *FW*...” But best in what way? Some sort of pragmatic appeal would be a natural thought here, starting with the very approximate idea that “the best moral *FW*” denotes whichever moral *FW* is most useful. As usual with pragmatic proposals, though, certain glaring questions jostle for attention and threaten to overturn the solution: Useful to whom? Useful in what way? And in this case there's the special question of “Useful when how grasped?”—in other words, is the best fiction that which would be most useful *if believed*, or that which would be most useful if the object of *make-believe*?

Let us remind ourselves what the cognitivist hermeneutic fictionalist is trying to achieve. He or she wants a theory according to which moral judgments stand a chance of being true, while remaining ontologically innocent with respect to problematic moral properties. Suppose that Ernie claims that Hitler was evil, while Bert (a well-known Nazi sympathizer) claims that Hitler was a moral hero. Error-theoretic worries threaten to render both speakers mistaken (along with speakers of every other first-order moral claim), a result from which most philosophers recoil. So Ernie and Bert's moral judgments are interpreted as prefixed with a tacit story operator. If one were tolerant of relativism, then one might consider appropriately relativized moral fictions, allowing both claims to be true: In Ernie's moral fiction Hitler is indeed evil; in Bert's moral fiction Hitler is indeed a moral hero. But this too is a result from

which most philosophers recoil. What most philosophers want (nice philosophers, that is, like us) is for Ernie's claim to come out as true and Bert's to come out as false. Therefore we—onlookers to their disagreement—will plump for a moral fiction that provides this result: In *our* best moral fiction, Hitler was evil and not a moral hero. We can interpret both Ernie's and Bert's claims as prefixed with the same operator concerning the same fictional world, thus providing an interpretation, sans problematic ontological commitment, according to which Ernie speaks truly and Bert speaks falsely. It would suit us very well if we could provide a non-arbitrary rationale for why our favorite moral fiction really is *the* best moral fiction, but even failing this the cognitivist fictionalist theory seems to have accomplished what was asked of it.

The noncognitivist hermeneutic fictionalist, by contrast, is not so obsessed with salvaging moral *truth*, but is nevertheless motivated to avoid attributing widespread error. Speakers whom we might be ordinarily tempted to see as making moral assertions and holding moral beliefs are interpreted as doing something else. Perhaps both Ernie's claim that Hitler was evil and Bert's claim that Hitler was a moral hero are equally false (for the kinds of reasons that impress the error theorist), but if neither of them is really *asserting* his claim (nor *believing* it), then neither can be charged with making an erroneous ontological commitment. Of course, acts of make-believe can be evaluated on other grounds. (The terms "pretense" and "make-believe" possibly do more harm than good in this context—with their implications of frivolity and superficiality—but I'll stick with them here.) Perhaps Ernie's type of make-believe is prudent or socially beneficial in some way, while Bert's is imprudent and socially injurious. So, as before, we, the onlookers to their moral disagreement, can take sides, and perhaps even have a non-arbitrary rationale for doing so. I will return to this matter later in this chapter.

### **But can it really be so easy?**

But can it really be so easy to side-step massive error simply by adding a tacit story operator or by interpreting people as "just pretending"? Should we even want to?

Ancient Romans were a superstitious bunch: divination, charms, omens, astrology, and necromancy were all widely accepted. Consider a representative claim: "An amulet can magically protect its bearer." Error looms, yet is easily avoided: We can interpret the Romans who made such claims as having asserted true sentences along the lines of "In FW so-and-so, an amulet can magically protect its bearer." Or we can interpret those speakers as having withheld assertoric force from their utterances of "An amulet can magically protect its bearer." Or we can interpret such utterances as expressing the cohortative sentence "Let's pretend that amulets can magically protect their bearers." I think it's safe to say, however, that *we feel no pressure to do any of these things*; indeed, we have no inclination at all to rescue the Romans from the falsehood and error of their superstitious ways. (Even if we were inclined to see such superstitions as serving some social good, we will classify it as a case of useful *falsehood*.)

Ancient Romans also by and large endorsed the geocentric view of the universe advocated by Plato and Aristotle and standardized by Ptolemy. This allowed them to employ an absolute notion of *motion*: things move relative to the Earth, which does not move—in other words, things move *period*. Consider a representative claim about motion: “Caesar moved his camp.” If the *movement* mentioned is taken to be absolute, then, because there is no such thing, the sentence is false. In this case, however, we feel less inclination to lumber the speaker with falsehood. Rather, we charitably interpret the predicate “...moved...,” which the Roman speaker and audience took to be an  $n$ -place predicate, as an  $n+1$ -place predicate—as tacitly relativized to a frame of reference. (The indeterminate reference to “ $n$ ” here is because “move” can be either a transitive or intransitive verb.) In other words, we interpret the speaker as using the same relativistic notion of motion that we use, and we may do this even while aware that the speaker might *deny* that this is what he or she meant (after all, he or she might be a committed Platonist about geocentricity). (See Harman and Thompson 1994: 4.)

Philosophers sometimes take it as a methodological principle that we should interpret the folk charitably, but the difference between these two cases reveals complications: in one we seem entirely comfortable about ascribing pervasive false assertions. (It would be a mistake to invoke Quine’s maxim of translation at this point, since this pertains to interpreting language *as a whole* and is entirely consistent with attributing massive error so long as the error is not inexplicable. See Boghossian and Velleman 1989: 97.) An obvious difference between the two cases is that in one but not the other *we* currently use a concept that is fairly obviously a close continuer to the ancients’ concept. Relative motion is not so very different from absolute motion in the sense that which concept is employed makes not a jot of practical difference in 99% of everyday cases; nearly all the time, we and the ancients employ the same frame of reference for our motion claims (namely, the Earth)—it’s just that we can recognize that this is a contingent choice while the ancients would have considered it mandatory. By contrast, we do not employ any concept that is a close continuer to the idea of the magical powers of amulets. Suppose that the closest we come is recognizing that some objects have sentimental value. But *having sentimental value* is simply not the same thing, and cannot be used for the same practical purposes, as *being magically protective*.

Charitably interpreting motion claims as involving a relativistic rather than absolute concept is one thing; charitably interpreting speakers as employing a tacit story operator, or as engaged in make-believe rather than belief, may be significantly different. For a start, either of the latter two expedients seems entirely too easy. Take any widespread apparently false belief that you like: the error of which the speakers apparently fall foul can be magicked away with a click of the fingers. Far from being a reasonable methodological principle of charity, this seems more like the manifestation of a disgraceful disregard for any epistemological standards. And it is, moreover, likely to lead to ruin, for what point is there in striving for truth when any error can be so easily ducked? When real-life natural disasters strike, rescuing *everyone* is a wonderful aspiration; but when it comes to epistemological methodology, the aim of rescuing *everyone* is itself a kind of disaster. If these devices

of hermeneutic fictionalism are going to be of any use, then, there will have to be some principled way of discriminating those error-threatened discourses for which the fictionalist solution is reasonable from those many discourses that are best interpreted as simply utterly mistaken.

Possible ways of making such a discrimination can be divided into those based on external considerations and those based on internal factors. Basing the distinction on external considerations means that some discourses warrant charitable interpretation (whereas others do not) in virtue of bearing some specifiable relational property to us (which the latter lacks). Earlier I speculated that we are motivated to interpret the ancients' talk of motion charitably because we continue to use a concept that is for most practical purposes indistinguishable from their use of absolute *motion*. If their talk did not bear this relation to our talk, then maybe we'd be willing to pronounce all their motion talk false. Alternatively, perhaps we find that some discourse is indispensable to our belief system, making us highly motivated to interpret it as non-erroneous, but we also recognize that it is deeply problematic when taken at face value, and thus we turn to fictionalist interpretations as a kind of last resort. If this discourse were not so indispensable to us, then maybe we'd be willing to pronounce it false.

Basing the distinction on internal factors means that some discourses warrant a fictionalist interpretation (whereas others do not) in virtue of having discernible intrinsic features (which the latter lack) indicating that the interpretation is actually reasonable. There are, after all, various ways in which the use of story operators or engaging in make-believe (rather than belief) can reveal themselves. To choose some simple illustrations: If a population of speakers had a tendency to consult a canonical fictional text before making any pronouncement on the topic of X, then this might encourage us to interpret their subsequent utterances about X as elliptical for "According to the fictional text..." Or if a population of speakers showed a tendency in serious contexts to back off from the claims that they make about X in everyday contexts, then we might be inclined to interpret their everyday discourse about X as a kind of make-believe. For example, people commonly talk and think about the sun rising, but if pressed they'll admit that this is false and in fact the observer's position on the Earth is rotating toward the sun. (Of course, if they were sensibly clear-headed relativists about motion, then they could maintain that the sun really does rise relative to the chosen frame of reference (see Jackson 2007), but this is not what most educated people do say—rather, when speaking carefully they'll deny that the sun really rises.) Ordinary talk of sunrises, then, appears amenable on internal grounds to a kind of fictionalist interpretation. (See Boghossian and Velleman 1989: 101; see also Van Inwagen 1990: 102-103.)

Were the decision to interpret moral discourse in a fictionalist manner based on the observation of internal features of the discourse calling for such an interpretation, then we would no longer be motivated by charity in particular. Rather, we would be in the altogether more secure position of responding to evidence. Basing one's decision on external factors, by comparison, introduces a rather unattractive kind of relativism into proceedings. Suppose we felt pressed to give an error-threatened discourse a



fictionalist interpretation as a last resort, because we judge the discourse indispensable to our conceptual scheme and find the prospect of its failure intolerable. If aliens with a different conceptual scheme were to face the same decision regarding the same discourse, they may find it more dispensable and therefore would lack our rationale for the fictionalist interpretation; they would plump for an error-theoretic interpretation of the discourse. If the matter depended entirely on external relational factors, then there would be no saying who is correct: us or the aliens. No further scrutiny of the discourse in question would expose evidence to resolve the disagreement.

Another difficulty for the externalist hermeneutic fictionalist is that it remains questionable just how charitable the fictionalist interpretation really is if it can be foisted upon speakers who would actually object. Suppose Amy says “Stealing is morally wrong” and we philosophers, knowing that the threat of an error theory looms, decide to interpret Amy as asserting something like “In the moral *FW*, stealing is morally wrong,” thereby construing her to be saying something true rather than false. But now what about Amy’s belief (and, one imagines, adamant declaration) that she is *not* using anything like a story operator (nor make-believing) when she engages in moral discussion? Presumably we’ll have to interpret *these* beliefs of Amy’s as false—so it’s not clear what favors we’d really be doing her. (See Friend 2008: 16.) By contrast, if the fictionalist interpretation is based on *internal* factors, then at least we would have some evidence to present to Amy to attempt to persuade her that the fictionalist interpretation of her moral discourse is reasonable. What kind of evidence might this be?

Let me focus on the noncognitivist brand of hermeneutic fictionalism here. One way of posing the challenge is as a response to what Matthew Chrisman describes as “a flatfooted phenomenological worry about moral fictionalism: it just doesn’t seem to me that I am operating under some pretense when I make a moral claim” (2008: 7). The fictionalist response can begin by emphasizing that it is not being claimed that our attitude toward morality is the same as that which dominates our engagement with familiar or childish fictions. The “make-believe” of which the fictionalist speaks need only bear *some resemblance* to more ordinary pretense—but precisely what kind of or degree of resemblance is vague and unspecified. It has, for example, often been objected to hermeneutic fictionalism that one of the hallmarks of pretense is a tendency to disengage with the fiction when the going gets tough (e.g., to abandon one’s captivation with the play when someone at the back of the theater shouts “Fire!”), but no such tendency is apparent in moral discourse (Chrisman 2008: 7; Cuneo 2014: 175). However, the kind of similarity that the hermeneutic fictionalist touts may not purport to preserve *that* feature. After all, it is doubtful that everything deserving of the name “pretense” exhibits this tendency. Lying is a kind of pretense—where the speaker pretends to believe something which he or she doesn’t believe—but sometimes people will not admit their lie in any circumstances (and sometimes have very good reason to refrain from doing so). Whatever kind of similarity the fictionalist does focus on, the really crucial thing, recall, is that the similarity be such that the ontological non-commitment characteristic of pretense is preserved.

The second thing the fictionalist can do is highlight the many other instances where we lack first person authority over what we are doing when we are speaking. To give a simple example: Many educated people insist that “data” is the plural of “datum” (adding that the widespread practice of using “data” as a singular term is a vulgar error). They are oblivious of the fact that *nobody* uses “data” consistently as a plural term (nobody says “One datum, two data, three data,...”); rather, it operates more like a mass noun. Of course, people can be pretty easily brought to see the error of their meta-linguistic beliefs in this case. So the third and most important thing the fictionalist needs to do is provide actual evidence that reveals the error of our meta-linguistic beliefs about the cognitive nature of moral language. (See discussion of “attitude-hermeneutic nominalism” in Rosen and Burgess 2005.) I have here put matters in terms of language, but matching points could be put, *mutatis mutandis*, about mental states. Not many people deny that we often lack first person authority regarding our own mental states.

Mark Kalderon (2005) sets out to do this by showing that the norms that govern moral discourse differ from those that govern assertion and belief. When one believes something, Kalderon claims, then upon encountering an epistemic peer who firmly disagrees, one has a “lax obligation” to examine one’s reasons for believing as one does. Kalderon calls this “noncomplacency.” However, the norms surrounding morality, he argues, permit complacency: we feel no embarrassment in steadfastly maintaining our moral views in the face of disagreement from epistemic peers. Were this argument to succeed, then we would have grounds for doubting that moral discourse is belief-expressing, but evidently more would need to be said to establish a similarity with fiction-talk sufficient to justify the label “fictionalism.” Kalderon himself offers little on this score, and claims that he means “pretense” only in a thin and non-explanatory sense (2008: 36). Kalderon’s argument is complex and I don’t propose to critically evaluate it here (I do so at some length in Joyce 2011); I mention it only to give some idea of the kind of thing the internalist hermeneutic fictionalist might say.

Wittgenstein offers some intriguing thoughts, with potential to develop into a fictionalist argument, when he notes that in moral discourse “we seem constantly to be using similes.” He continues:

But a simile must be the simile for something. And if I can describe a fact by means of a simile I must also be able to drop the simile and to describe the facts without it. Now in our case [i.e., ethics] as soon as we try to drop the simile and simply to state the facts which stand behind it, we find that there are no such facts. And so, what at first appeared to be simile now seems to be mere nonsense. (Wittgenstein 1965: 10)

Wittgenstein seems to be somewhat misusing the term “simile,” for an explicit comparative sentence of the form “X is like Y” does not stand for anything (except in the trivial sense that applies to all sentences). (Wittgenstein’s careless use of the term “simile” is noted in Erden 2012.) But moments earlier he talks of terms being used “as similes or allegorically” (1965: 9), which makes clearer that he has in mind terms

that function to convey symbolic meaning, but for which the hidden meanings turn out in these cases to be absent. Since allegories are devices prevalent in fiction, where false images, sentences, and narratives are presented without ontological commitment to the manifest content of the allegory, it would not be unreasonable to classify Wittgenstein's view as a form of hermeneutic fictionalism. (It's worth mentioning that Wittgenstein's contention that the allegorical language is empty is superfluous to this classification; even if moral discourse consisted of allegories that successfully refer to specifiable truths, the fictionalist interpretation would be in order.)

After making this intriguing claim about moral language consisting of similes [sic], however, Wittgenstein pretty much leaves his audience to seek their own evidence that this might be true. I shall not attempt to add anything on that score, but simply use this as an opportunity to illustrate that the empirical nature of the internalist fictionalist enterprise is a mixed blessing. Sensitivity to empirical evidence opens the possibility of hermeneutic fictionalism's being well-grounded and confirmed; the problem is that the evidence seems thin, to say the least. Fictionalists can help themselves to whatever evidence noncognitivists have been able to muster for the hypothesis that moral discourse is not belief-expressing (this evidence is far from conclusive, but at least there's a decades-long tradition with several well-worn paths to explore); what is currently lacking is a well-developed program of seeking evidence for the positive part of the fictionalist hypothesis: that interesting similarities between moral discourse and fiction-talk exist.

Stephen Yablo (2000; 2005) has argued for something like hermeneutic fictionalism regarding numbers partly on the basis of observations concerning both ordinary people's and mathematicians' number discourse, such as their apparent indifference to, and impatience with, questions regarding the ontology of numbers. (If one doesn't care whether numbers exist, then presumably one's mathematical discourse is not ontologically committed to them.) It is not obvious that analogous arguments would be any less plausible if pressed into the service of moral fictionalism. (See Hussain 2004 for discussion of how Yablo's ideas might apply to the moral case.)

### **Revolutionary moral fictionalism**

In some ways, the revolutionary fictionalist has it easier than the hermeneutic fictionalist. There is no need to locate evidence that fictionalism is *true*, for the revolutionary fictionalist doesn't claim that mark of distinction for the theory. Rather, the revolutionary fictionalist needs to sell the theory as *good advice*. (Of course, first the truth of error theory needs to be established, but, barring one exception later, we'll bracket off those arguments in this chapter.) Regarding the content of the advice, revolutionary fictionalists have the same menu of options as their hermeneutic relatives. One option is to revise one's moral utterances so they are prefixed with a (usually tacit) story operator. Another option is to alter one's attitude toward morality from belief to something akin to make-believe. As with advice in general, the recommendation must be based on some sort of cost-benefit analysis. "Morality is

useful when believed,” the fictionalist reasons, “so eliminating it will, *ceteris paribus*, be costly. Perhaps some of the benefits of morality which we stand to lose can be recouped by taking a fictionalist stance toward morality.”

The first thing to notice is that the fictionalist need not claim that taking a fictional stance toward morality is *just as good* (in pragmatic terms) as believing it. Perhaps the fictional stance is, in practical terms, a sorry second best to a believed morality. But sincere belief is out of the running at this stage of the argument. The fictionalist’s relevant competitor is the eliminativist, who advocates abolishing moral discourse altogether (which is not, of course, the same as declaring a moratorium on even uttering any moral terms). The second thing to notice is that the revolutionary fictionalist’s reasoning (as just sketched) clumsily obscures the relativism inherent in any advice-giving. Whenever a philosopher claims that something is useful, one’s immediate thought should be “To whom?” We may have knee-jerk intuitions that morality is broadly useful, but is it really useful to *everyone*? (Was it useful to the men dead in Flanders fields, brought to their sorry ends by moralistic propaganda and their sense of duty?) Perhaps taking a fictional stance toward morality will recoup costs for one person but not for another. Even the best advice is unlikely to be good for anyone in any circumstances. In light of this, the revolutionary fictionalist should be permitted a degree of modesty and a dose of vagueness: The position is reasonable if it’s good advice generally for most people.

A particular aspect of this relativism is worth highlighting. One fictionalist proposal is that we give up asserting moral claims but rather use them to perform another speech act—one modeled on some aspect of fiction-talk. But speech acts occur only against a background of conventions shared by a speaker and her audience; a person cannot unilaterally decide that she isn’t asserting the sentence *S* if she fails to signal this to her audience, all of whom take her to be asserting *S*. Given this, it is problematic to consider the fictionalist’s advice as directed toward *individuals*, for it is not clear that it could even be coherent advice for an individual. (It would be like advising someone to become a rugby team.) Continuing to bear in mind the modesty and vagueness, then, it is best to consider the fictionalist’s advice as directed toward *groups* of speakers.

There are different ways of understanding the claim that *X* is useful to a group, even before we get to more specific questions raised by replacing “*X*” with “morality.” Let us suppose that we settle on one such way. If a group is motivated to adopt morality as a fiction because doing so is useful (in the manner settled upon), then when faced with the choice of *which* moral fiction to adopt (from an infinite range of possibilities), the answer is simply “The most useful one.” It is important to remember that the fiction is being maintained for practical purposes; it is entirely possible that a group might adopt the wrong moral fiction.

Thus the (false) moral claims are grounded in (true) practical claims. This is often the case with fiction-centered activity. Kendall Walton (1993) distinguishes between *content-oriented* make-believe and *prop-oriented* make-believe. In the former, the players’ real interest lies in the fictional world, and the props are but tools in the service of that end. (An up-turned couch is a ship, a stick is a gun, and so on.) In the

latter, it is the prop itself that is really of interest, and the make-believe game is a way of revealing its features. One of Walton's examples involves telling someone the location of the town of Crotona by saying "Imagine Italy is a boot; then Crotona is on the arch of the boot." (And if that's insufficient, I might even remove one of my stiletto boots and point: "Crotona is *here*.") Metaphor is a central example of a kind of pretense where the focus is not on the fiction per se, but on truths that it reveals. Someone who says "Fred is a two-faced snake" is trying to tell us something true, albeit somewhat indefinite, *about Fred*. (Versions of fictionalism that emphasize that the fiction is a means of getting at important truths include Yablo's views on mathematical discourse (2000, 2005) and Arnon Levy's views on discourse about biological information (2011).)

The fiction of morality, then, need not be just a wild and whimsical falsehood in which we indulge for practical benefit. The falsehoods of morality can be ways of drawing attention to truths about what will not be tolerated, what is most valued, what will be harmful, etc. Although pointing to the arch of a boot and saying "Crotona is here" is obviously false, saying this while pointing to *the toe* of the boot would be worse. (I am tempted to say "doubly false.") Analogously, the revolutionary fictionalist may maintain that while saying "Hitting babies is morally wrong" is false, saying "Hitting babies is morally permissible" would be worse.

One immediate question for the fictionalist is this: If one can speak truly about what will not be tolerated, what is most valued, what will be harmful, etc., then hasn't the error theory evaporated? Why not just accept that these truths *are* the moral truths (or at least the base upon which the moral truths supervene)? Clearly, however, this is not something that is generally true of metaphors. Saying that Fred is a two-faced snake does not invite the rejoinder "But if by this means you're able to make reference to traits that Fred truly has (being deceitful, sneaky, etc.), then surely it's literally true that Fred is a two-faced snake." In the case of morality, we are assuming that successful arguments for the error-theoretic position are already on the table, and these arguments will have shown that moral normativity has some special and problematic qualities that other norms (pertaining to what will be tolerated, etc.) do not have. It is not in the remit of this chapter to go into those arguments, and doing so would be time-consuming. Suffice it to say that if our discussion of revolutionary fictionalism begins "Suppose for the sake of argument that arguments for a moral error theory were successful..." then we have already made the supposition (if only *arguendo*) that the revolutionary fictionalist has the resources to rebut this criticism.

If morality is a fiction that tracks truths that are important to us, then this explains why we might care about a fiction, and why we might care more about one moral fiction than other possible moral fictions. What it doesn't explain is why we should need or want the fiction at all. Why not just talk in terms of literal truths? Why not, in other words, embrace eliminativism?

Well, why do we use metaphors at all? Why not just talk in terms of literal truths all the time? "Metaphor," answers Dick Moran, "does appear to have a force that goes beyond agreement with what it asserts" (1989: 91).

To call someone a tail-wagging lapdog of privilege is not simply to make an assertion of his enthusiastic submissiveness. ... [T]he comprehension of the metaphor involves seeing this person as a lapdog, and in some detail, experiencing his dogginess. This is what a successful metaphor pulls off, and this image-making quality is what lies behind both the force and the unparaphrasability of poetic metaphor. (1989: 90)

And here is Stephen Yablo discussing Walton:

A certain kind of make-believe game, Walton says, can be “useful for articulating, remembering, and communicating facts” about aspects of the game-independent world. He might have added that make-believe games can make it easier to reason about such facts, to systematize them, to visualize them, to spot connections with other facts, and to evaluate potential lines of research. (Yablo 1996: 279)

I will finish the discussion of revolutionary fictionalism by tying these thoughts about the power of metaphor back to the case of morality. First, though, I will need to quickly sketch an argument that might motivate the revolutionary fictionalist’s error-theoretic leanings in the first place. (Bear in mind that a moral error theory might be based on a different argument entirely.)

Suppose we come to see that the only correct view of human ends is a broadly Humean picture: our ends depend on our contingent desires (though possibly those held under various kinds of idealization). Suppose also that we come to see that an essential component of moral discourse is a reference to non-Humean ends: Morality often deals in matters of what we “have to do” whether we like it or not; it is imbued with an inescapability that the Humean view cannot underwrite. This would be the basis of a moral error theory (see Mackie 1977; Joyce 2001). Suppose further, however, that *thinking in terms of* the Humean picture tends to reduce the probability of our actually achieving our ends. We are more likely to achieve our real Humean ends if we picture those ends in non-Humean terms: if we think of them as things that we *must* pursue whether we like it or not. Why might this be? Perhaps deliberating in terms of ends that brook no discussion bolsters our motivation to pursue them. By comparison, thinking of those ends in Humean terms allows them to wear their contingency on their sleeves. The thought “If I didn’t desire such-and-such, then I wouldn’t have the reasons that in fact I do have” can be a dangerous one; it opens the door to self-sabotaging rationalizations like “But, really, how much *do* I desire such-and-such?” We often succumb to temptations that we later regret—irrational lures that disrupt our ability to pursue our real ends. Perhaps a firmer habit of thinking in terms of *moral* norms—that is, norms that demand compliance irrespective of our desires—would help us achieve our goals. This would be an ironic twist of human psychology, to be sure, but it seems not entirely implausible that it is the bind in which we find ourselves. (See discussion of “conversation stoppers” in Dennett 1995.)

The fictionalist response to this bind is to recommend that we exploit the foibles of our own psychology by cultivating non-Humean thinking as an expedient for tracking, and better motivating the pursuit of, our actual Humean ends. Like a metaphor’s ability to draw attention to truths that might otherwise evade simple

description (or regarding which, at least, the non-metaphorical description might be cumbersome), moral language, false though it is (in the error theorist's opinion), encourages speakers and their audience to see the world in a certain way that might otherwise evade simple description and without which they might be left susceptible to forms of self-subversion. If we were all fully informed, fully reflective, rational, clear-headed, and strong-willed, then perhaps we would have no need of any moral fiction. Revolutionary moral fictionalism depends on the contingent fact that we tend to fall well short of satisfying this list of admirable qualities.

#### REFERENCES:

- Boghossian, P. and Velleman, J. 1989. "Colour as a secondary quality." *Mind* 98: 81-103.
- Chrisman, M. 2008. "A dilemma for moral fictionalism." *Philosophical Books* 49: 4-13.
- Dennett, D. 1995. *Darwin's Dangerous Idea*. NY: Simon & Schuster.
- Eklund, M. 2015. "Fictionalism." In E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL = < <http://plato.stanford.edu/entries/fictionalism>>.
- Erden, Y. 2012. "Wittgenstein on simile as the 'best thing in philosophy.'" *Philosophical Investigations* 35: 127-137.
- Friend, S. 2008. "Hermeneutic moral fictionalism as an anti-realist strategy." *Philosophical Books* 49: 14-22.
- Harman, G & Thomson J.J. 1996. *Moral Relativism and Moral Objectivity*. Cambridge, MA.: Blackwell.
- Hussain, N. 2004. "The return of moral fictionalism." *Philosophical Perspectives* 18: 149-187.
- Jackson, B. 2007. "Truth vs. pretense in discourse about motion (or, why the sun really does rise)." *Noûs* 41: 298-317.
- Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, R. 2011. "Review essay: Mark Kalderon's *Moral Fictionalism*." *Philosophy and Phenomenological Research* 85: 161-173.
- Kalderon, M. 2005. *Moral Fictionalism*, Oxford: Oxford University Press.
- Kalderon, M. 2008. "The trouble with terminology." *Philosophical Books* 49: 33-41.
- Levy, A. 2011. "Information in biology: A fictionalist account." *Noûs* 45: 640-657.
- Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin.
- Moran, R. 1989. "Seeing and believing: Metaphor, image, and force." *Critical Inquiry* 16: 87-112.
- Olson, J. 2014. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.
- Proudfoot, D. 2006. "Possible worlds semantics and fiction." *Journal of Philosophical Logic* 35: 9-40.

- Rosen, G. 1990. "Modal fictionalism." *Mind* 99: 327-354.
- Rosen, G. and Burgess, J.P. 2005. "Nominalism reconsidered." In S. Shapiro (ed.), *The Oxford Handbook of Philosophy and Mathematics and Logic*. Oxford: Oxford University Press: 515- 535.
- Van Inwagen, P. 1990. *Material Beings*. Ithaca: Cornell University Press.
- Vision, G. 1994. "Fiction and fictionalist reductions." *Pacific Philosophical Quarterly* 74: 150-174.
- Walton, K. 1993. "Metaphor and prop oriented make-believe." *European Journal of Philosophy* 1: 39-57.
- Wittgenstein, L. 1965. "Lecture on ethics." *Philosophical Review* 74: 3-12.
- Woodward, R. 2011. "Is modal fictionalism artificial?" *Pacific Philosophical Quarterly* 92: 535–550.
- Yablo, S. 1996. "How in the world?" *Philosophical Topics* 24: 255-286.
- Yablo, S. 2000. "A paradox of existence." In T. Hofweber and A. Everett (eds.), *Empty Names, Fiction and the Puzzles of Non-existence*. Stanford, CA, CSLI: 275-312.
- Yablo, S. 2005. "The myth of the seven." In M. Kalderon, (ed.) *Fictionalism in Metaphysics*. Oxford: Clarendon Press: 88-115.