Moral fictionalism: How to have your cake and eat it too
Richard Joyce

1. Introduction

There are good reasons for thinking that moral discourse is hopelessly broken. Engaging sincerely in moral talk seems to commit speakers to properties like moral obligation, having moral rights, blameworthiness, virtue and vice, etc., and there are, on reflection, significant grounds (metaphysical, epistemological, and empirical) for doubting that the world we inhabit contains anything like these properties. Discovering this mismatch between two theses—a conceptual thesis (concerning what our moral discourse commits us to) and an ontological thesis (concerning what the world is like)—leads to the error-theoretic result that none of our moral judgments are true.[1] (See Mackie 1977; Joyce 2001; Olson 2014.)

Not many people accept an all-out moral error theory, though partial moral error theories are more familiar. The notion of *sin*, for example, which used to be fairly central to moral discourse (in the West, at least), has dropped from many people's normative conceptual scheme, meaning that even those who have no sympathy with radical moral skepticism may well, upon reflection, consider themselves to be error theorists about sin. Much the same might be true of *evil*. Needless to say, there are still many people who believe in sin and evil; my aim is just to draw attention to the familiarity of this kind of doubt (recognizable even to those who don't harbor the doubt) and thus to conceptualize all-out moral error theory as a familiar kind of doubt writ large: extending to all moral concepts.

An all-out moral error theory accuses widespread and customary ways of thinking and talking of a committing a massive mistake, and is to that extent a counter-intuitive position which may be safely predicted to meet with considerable heartfelt resistance. But the resistance is born not simply of the fact that people are generally opposed to admitting that they (and probably everyone they know) are wrong. It is also born of the fact that there is a certain anxiety surrounding what might *happen* if the moral error theory is true—or, at least, if it came to be widely believed to be true. Morality is the bedrock of a civilized society (the thought goes), and any theory that comes along claiming that it's all baloney is therefore dangerous and its advocates pernicious.

On the face of it, this chapter (like this whole volume) is not about arguments for and against the error theoretic position; it is about what happens next, were that position to be adopted. But dig a little deeper and the distinction between these two questions is not so clear-cut. If resistance to the error theory is due partly to worries

---

[1] Whether we conclude that they are all false, or neither true nor false (as Howard Sobel thinks: this volume) is a nice distinction that, perhaps, doesn't matter terribly much. Another nice distinction to be put aside on this occasion is whether the failure of moral discourse is a contingent or necessary one.

about what might *happen* if it were widely adopted, then addressing the "what happens next?" question becomes a prerequisite for deciding whether and/or how strenuously to resist the theory. (See Lutz 2014: 370.) If a persuasive case were to be made that we could adopt the error theoretic position and civilization would not collapse—that life would go on as before, or even go on *better*!—then the opposition to the theory might diminish, or at the very least lose some of its determination.

The three positions that I shall critically assess in this chapter are abolitionism, conservationism, and fictionalism, and my general goal is to speak in favor of the last. Each of these positions is a competing answer to the question "If we were to adopt a moral error theory, then what should we do with our moral concepts?" In a nutshell, the abolitionist answers "By and large, stop using them," the conservationist answers "Keep using them, including in moral beliefs," and the fictionalist answers "Keep using them, but not in moral beliefs." As I say, which answer is to be preferred should be of interest not just to those who are tempted by or convinced of the moral error theorist's arguments, since the answer may well inform whether to be tempted or convinced in the first place.

In section 3 I shall compare conservationism and fictionalism, and in section 4 I shall compare abolitionism and fictionalism. First, in section 2, I shall sketch out the three options in a bit more detail.

2. Abolitionism, conservationism, and fictionalism

If we were to adopt a moral error theory, then what should we do with our moral concepts? Clearly, the "should" cannot be interpreted as a moral "should," and, more generally, must survive unscathed the error theorist's arguments against morality. The natural thought (though it is by no means a mandatory one) is that it is to be understood in broadly Humean terms: as some kind of function of individuals' desires (desires which need not be selfish). Since we are talking about a plural "we," then the matter can be broken into two steps: what an individual should do is a function of her desires, and what a group should do is a function of what the individuals comprising the group should do. Both steps allow various options (consonant with Humeanism), but I'm happy to leave matters open here. The important point is that whatever considerations have led the error theorist to doubt moral normativity are assumed not to apply to these Humean group norms, and a corollary of this assumption is that moral normativity essentially has a quality or qualities that Humeanism cannot satisfy. (These are contentious assumptions, to be sure, and well worth arguing over, but for current purposes it's a background supposition that those arguments have already fallen in the error theorist's favor.)

It's fair to say that, generally, when we have in the past become error theorists about something then abolitionism is the usual response.[2] Of course, we don't entirely

---

[2] In my 2001 book I used the term "abolitionism" (Joyce 2001), but subsequently tended to prefer the label "eliminativism" for this theory (see Joyce 2013, 2017). I harbor a small worry that "abolitionism" carries some historical baggage that could skew the debate (after all, in the 18th and 19th centuries the

refrain from using words like "unicorn," "phlogiston," and "witch" (any more than the word "God" never escapes the lips of an atheist), but we do stop making assertions that commit us to the existence of these entities. The abolitionist with whom we are concerned may be interpreted as *doubly* opposed to a subject matter: she is opposed to an ontology (she doesn't believe in witches, say), and she is opposed to a language (she recommends that we by and large stop speaking of witches).[3] The assumption that the second form of opposition follows from the first may appear so obvious that arguing for the first (that is, arguing for an error theory) will seem sufficient for establishing the second. But the assumption is mistaken, and the second form of opposition requires independent argument. An ontology should be opposed if one comes to see it as probably *false* (on either empirical or *a priori* grounds), but one should oppose a use of language if one comes to see it as *inadvisable*. To reject moral language, then, the abolitionist must argue that it is inadvisable, and this requires an analysis of its costs and benefits. Not only is establishing a moral error theory insufficient for establishing abolitionism, nor is it necessary, for even someone who believes in moral facts can nevertheless decide that moral discourse is ill-advised on pragmatic grounds (see Campbell 2014; Ingram 2015).

Nobody, of course, would claim that making moral judgments is *always* harmful, and nobody would maintain that doing so is always beneficial. Making a moral judgment might be beneficial to the judge but harmful to others, or vice versa. Making a moral judgment might be useful to the judge on Monday but harmful to the judge on Friday. Whether making a moral judgment is harmful or beneficial may also be frequency-sensitive: dependent on whether others (and how many others) are also disposed to do so. Some of these complications can be put aside if it is stipulated that we are talking about a *group* (as the question posed at the start of this section presumes). But other complications remain. There is, of course, the question of which function from members' goals to group goals should be utilized in order to allow us to speak in Humean terms of "what the group should do." Yet even if this were settled, it remains exceedingly implausible that making moral judgments will *always* be harmful to a group regardless of how it is situated, and so we shouldn't burden the abolitionist with any such far-fetched claim. Rather, the abolitionist's claim may be focused on *typical* human groups—that is, where resources are finite, where interests often come into conflict, where mutual benefits can be gained through cooperation, and so forth, and also where "typical" human psychology (cognitive skills, emotions, epistemic limits, etc.) is in play. If it turns out that the notion of "typicality" here is insupportable, then the abolitionist can always ground the claim in actuality: as a view about how costly morality is and has been for *us*. Several contributors to this volume—Garner, Hinckfuss, Marks, and Yaouzis—set out to highlight the negative

---

abolitionists were the *good guys*, right?). But the tide seems to be turning toward the label "abolitionism," and I shan't put up a fight.

[3] In fact, the abolitionist may be interpreted as *triply* opposed: she is also opposed to a way of *thinking*. For the sake of brevity, on this occasion I'm happy to fudge together language and thought. (The term "moral judgment" is quite useful in this regard.)

impact of moral judgments, either generally or with regard to specific topics of discourse—often focusing on the costs and benefits for *us now*, though also frequently discussing historical scenarios (see also Isserow, this volume).

Recall, however, that the kind of abolitionist in whom we are interested doesn't simply claim that moral discourse and thinking is inadvisable; she is also an error theorist, arguing that moral discourse and thinking is untrue. Thus it is not really *us* on whom her cost-benefit analysis should focus, but rather some of our counterparts: that is, us *if we came to accept a moral error theory* (which, of course, we haven't, and nor are likely to in the foreseeable future). This potentially makes a significant difference, since the costs and benefits of morality for a group *that rejects error theory* may be different from the costs and benefits of morality for a group *that accepts error theory*.

At this point one could be forgiven for wondering whether the abolitionist has rather wasted her energies discussing current and historical harms caused by moral thinking, for these are all harms that befall/befell us *when morality is believed*. And surely (the thinking goes) if we were to accept a moral error theory then the option of believing morality would evaporate, and so those types of harm may no longer be pertinent.

This is where the proponent of the second option, conservationism, enters the picture. The conservationist insists that our becoming error theorists would *not* cause the option of believing in morality to evaporate. The conservationist accepts that morality is often harmful (both to the judge and/or to others), but thinks that the abolitionist has undersold the benefits it brings: on balance, the benefits outweigh the costs (for typically situated groups of humans). Yet these benefits, the conservationist observes, have historically depended on morality's being *believed*. And therefore the conservationist simply recommends that we carry on believing in morality as far as everyday decisions in everyday life go. We may *disbelieve* morality when doing metaethics (after all, the conservationist believes that the moral error theorist is correct, and when discussing metaethics will assert this), but the rest of the time— which is most of the time—the advisable course is that we should just carry on with our old moral discourse and old moral beliefs, both of which commit us (erroneously) to properties like moral obligation, having moral rights, blameworthiness, virtue and vice, etc.

It should be noticed that the world about which the conservationist conducts his cost-benefit analysis should not really be the actual world, for much the same reason as before: the costs and benefits of morality for a group *that rejects error theory entirely* (which is us now and in the past) may be different from the costs and benefits of morality for a group *that accepts error theory when doing metaethics but rejects it when engaged in everyday life*. For example, the latter group must "compartmentalize" their beliefs in a way that the former group need not, and this very practice may incur additional costs. (Indeed, as we'll see later, this is likely to be a central argument that the abolitionist will employ against the conservationist.) So conservationism, like abolitionism, relies on the truth of a counterfactual. That is not to say that the abolitionist and the conservationist have no business wrangling over

the net costs and benefits of morality in the actual world, but the debate must be handled with care: the cost-benefit analysis of morality in the actual world is relevant only to the extent that it casts light on the cost-benefit analysis of morality in another possible world.

The fictionalist, our third contender, agrees in part with the abolitionist and in part with the conservationist. With the abolitionist the fictionalist agrees that we mustn't commit ourselves to the existence of things in which we disbelieve. She (the fictionalist) also agrees with the abolitionist that the conservationist's recommendation involves doxastic practices that are likely to lead to deleterious consequences. The conservationist is, after all, recommending that we knowingly cultivate inconsistent beliefs, which is, arguably, a recommendation of irrationality. The fictionalist and the abolitionist together recoil at such a violation of epistemic norms.

With the conservationist the fictionalist agrees that morality is and has been, on balance, useful (for typically situated groups). The fictionalist's distinctive claim is that morality can to some extent remain useful even if it is not believed. The fictionalist looks to the way familiar fictions operate as a model for how we can usefully talk about topics without committing ourselves to the entities under discussion. Just as we can discuss Sherlock Holmes without committing ourselves to his existence, treat a stick as if it's a sword without committing ourselves to its being so, or complain that an enemy is a "spineless snake" without committing ourselves to unorthodox views on zoological taxonomy—so too (the fictionalist maintains) our error theorist counterparts could talk about promise-keeping as morally obligatory (say) without committing themselves to the existence of moral obligations.

The fictionalist in whom we are interested does not claim that our current moral discourse is anything like a fiction; rather, she thinks that were we to realize the truth of error theory then we should *change* our moral discourse, to make it in some sense similar to fiction, so as to remove commitment to entities in which we disbelieve.[4] But, of the infinity of possible moral fictions, *which* one should be chosen: the one in which promise-keeping is obligatory regardless of consequences? the one in which promise-keeping is valued but overridable? in which promise-keeping is supererogatory? in which it is neutral? frowned upon? evil? The answer is simply that a group should choose whichever moral fiction best satisfies their Humean ends.

I have acknowledged that there are many challenges to be overcome surrounding this type of answer, but my point here is that these are challenges that arise simply from the way we have posed the question "If we were to adopt a moral error theory, then what should we do with our moral concepts?"—they arise for fictionalist, abolitionist, and conservationist alike (and, indeed, anyone else who wants to offer an answer)—and so it is not a special problem for the fictionalist at this point in the dialectic. I also acknowledge that it is possible that one might hold that Humeanism

---

[4] The fictionalist under discussion in this chapter is the *revolutionary* fictionalist. Someone who thinks, by contrast, that our current moral discourse is already somewhat like a fiction is the *hermeneutic* fictionalist. For advocacy of the latter view, see Kalderon 2005. For further discussion, see Joyce 2017.

works for individual practical rationality but (for some reason) isn't feasible for groups. Very well, then; in that case we shall have to reformulate our original question so as to concern what an *individual* should do if he or she adopts a moral error theory. Our three contending theories can still compete as answers to that question, though we would face the additional complication of relativism: that one individual's interests might best be served by conservationism, another individual's interests might best be served by fictionalism (and one moral fiction rather than another), and so on. It may come to this; but for now let's be optimistic Humeans and assume that the original question can stand.[5]

Think of the moral fiction as an overlay on the group's actual Humean goals—an erroneous overlay that allows the group to better achieve its Humean goals. Of the types of fictional discourse that are familiar to us, perhaps the best model for the fictionalist is *metaphor*. When I believe that someone (call him "Jake") is dishonest and cowardly then I could just describe him as such, but I also have the option of speaking figuratively: I could say "Jake is a spineless snake!" If you ask me whether I really believe that Jake is a spineless snake, then, caught up in my annoyance, I'm likely to answer "Yes!" But if you press me in the right way—"So, wait, you believe that there are *literally* invertebrate reptiles, and that this apparent human, Jake, is *literally* one of them??"—then I'm likely to back off: "No, of course I don't really think *that*, but you're missing the point!" There are facts about Jake that I want to convey—his dishonesty and cowardice—but by saying something that is literally false I draw attention to those facts more evocatively and dramatically, which might serve my goals better than straight-talk.

Let's vary the example slightly, and suppose that I described Jake as "an *immoral* spineless snake." The fictionalist pictures a world where what goes for "spineless snake" goes for "immoral." In this world, if you ask me whether I really believe in immorality—where immorality involves (let's suppose) things that mustn't be done regardless of one's goals (i.e., non-Humean norms)—then at some point I'll back off: "No, of course I don't really think *that*, but you're missing the point!" There are facts about Jake that I will want to convey—this time concerning his frustrating of the group's Humean goals—but by saying something that is literally false I draw attention to those facts more evocatively and dramatically, which might serve my goals (or the group's goals) better than straight-talk.[6]

I will return to this fictionalist model of metaphor repeatedly in the following critical discussion of these three theoretical options.

---

[5] In fact, of course, we already face the complication of relativism, since what one group ought to do in this respect may differ substantially from what another group should do. I don't think we should fear this relativism. The abolitionist, conservationist, and fictionalist are all in the business of offering advice, and good advice rarely, if ever, holds universally. Consider the excellent advice "Eat your greens."

[6] My original example here involved calling Jake "an immoral dickhead"—but then I worried that the ensuing explanation (of what would be involved in someone's being literally "a dickhead") might be distracting.

3. Fictionalism vs. conservationism

In order to assess the debate between the fictionalist and the conservationist, we first need a working hypothesis (at the very least) of the benefits that morality has brought. Limitations of space restrict me to focusing on just one plausible way that moral thinking is generally useful: by strengthening our motivation to act. Judging an act to be "something that must be done whether I like it or not" (i.e., as required in non-Humean terms) may strengthen one's resolve to perform it; in certain contexts it may even be motivationally superior to the thought "this action would satisfy my desires." After all (the thinking goes), the latter judgment seems to invite inner negotiation: "But how much do I really desire such-and-such?"—allowing for all-too-familiar rationalizations that ultimately amount to self-sabotage. By contrast, thinking "I just *must* do it" works to shut-down inner debate (though by no means guaranteeing that the action in question will be performed).

Of course, having at one's disposal psychological mechanisms that strengthen motivation isn't necessarily useful—it depends on which motivations are being strengthened! Where such mechanisms might be most useful is when they promote faint and intangible desires that are at risk of being overwhelmed by more immediate and concrete desires (on the assumption that the satisfaction of the former desires really is more useful). An obvious commonplace example of this pattern is where desires for the benefits of living in a cooperative society are at risk of being overwhelmed by short-sighted selfish temptations. And this, it seems to me, is a quite plausible and intuitive sketch of one basic way in which moral thinking is generally useful: it acts as a bulwark against nearsighted self-centeredness, strengthening our motivation to act cooperatively by providing no-nonsense (non-Humean) imperatives in favor of doing so. Using this hypothesis, we can turn to the debate between the conservationist and the fictionalist.

The conservationist worries that merely "make-believing" that morality is true is likely to strip it of its usefulness. The conservationist allows that the moral thought "I must cooperate, whether I like it or not" may well be useful in the hypothesized manner, but surely only when it is an item of *belief*. The conservationist sees the fictionalist as striving to avoid falling into the habit of moral belief by practicing "cognitive self-surveillance" and having to occasionally remind herself of the truth of error theory "in order to prevent slipping into holding real moral beliefs and making genuine moral assertions" (Eriksson & Olson, this volume).[7] "But this reveals a deep practical tension in moral fictionalism, for it also seems that in order for moral precommitments to be effective in bolstering self-control, beliefs to the effect that morality is fiction need to be suppressed or silenced" (Olson 2011: 197).

---

[7] Terence Cuneo and Sean Christy make a similar point: "To operate in this way requires not only that the folk reliably keep critical and ordinary contexts distinct, but also that they exercise remarkable discipline and imagination when in ordinary contexts, governing their belief-forming faculties in such a way that they do not produce moral beliefs. Perhaps this could be accomplished in some way. But we imagine that for many this will prove psychologically very difficult" (Cuneo & Christy 2011: 98).

In place of this tension, the conservationist recommends that we give up the fight and surrender ourselves to moral *belief*. That way, there is no need for constant "self-surveillance" and no risk that moral judgment will lose the motivation-strengthening qualities that it enjoyed before we became error theorists. It may be true that the person simultaneously believes in the moral error theory, but this metaethical belief can be utterly suppressed in everyday contexts, the conservationist thinks, and that's all to the good as far as the usefulness of morality goes. By contrast, there is something problematic in the fictionalist "utterly suppressing" her skeptical metaethical beliefs, for then the suspicion arises that she will in fact have slipped into moral *belief*, thus betraying her fictionalism.[8]

Reflection on how familiar metaphors work should reveal, I believe, that these anti-fictionalist worries are misplaced. When I call Jake "a spineless snake," surely I do not need to practice any self-surveillance to ensure that I avoid inadvertently slipping into believing that he literally is an invertebrate reptile.

Or consider the metaphor "I love you with all my heart." The association of the heart with love goes back historically to a time when it was believed that the heart is the seat of human emotion (as Aristotle thought); so perhaps in the time of the ancient Greeks the sentence could be intended literally. But now we know better, and when one uses the phrase there remains no temptation whatsoever to inadvertently believe that one's love really is the product of cardiac activity. Everyone now knows that the speaker of the phrase is employing a falsehood to convey an important truth. The speaker could, of course, eschew the falsehood and talk in literal terms, but there may be all sorts of pragmatic reasons for preferring not to. Declaring love is one context where speaking in precise and unembroidered language is especially likely to be ill-advised.

Consider a third example: of someone accusing me of being a pig at dinner, to which I indignantly respond "I was *not* a pig at dinner!" In this case, the sentence I've uttered is actually *true* (for at dinner I was not, nor have I ever been, a pig), and yet the sentence still employs a metaphor. But even here, where I've uttered something that is literally true, no one (least of all me) is likely to interpret me as having asserted the obvious fact that I was not, during dinnertime, temporarily a member of the genus *Sus*. I have *not* asserted this fact (although it is a fact, and, had I wanted to, I could have used the very same sentence to assert it); rather, I've used the sentence to convey some different information, concerning my not violating the demands of etiquette at dinner. In short, we already have entrenched and well-developed skills at saying and thinking false things in order to convey important truths (or, in the third example, saying an obviously true thing in order to convey a very different and less obvious truth), and the conservationist underestimates our capacity to do this smoothly and without confusion.

---

[8] I use the label "fictionalist" to denote both a person who advocates fictionalism and a person of whom fictionalism provides an (allegedly) accurate psychological description. Troubled by this ambiguity, Cuneo and Christy (2011) introduce the term "fictioneer" for the latter. But I find the ambiguity innocuous.

The other component of the conservationist's doubt about fictionalism is the worry about how an act of make-believe could be "effective in bolstering self-control" unless the knowledge that it is a fiction is "suppressed or silenced." But I think again that the example of metaphor fairly easily answers this doubt. Consider once more the earnest declaration: "I love you with all my heart." This utterance, and the associated thoughts, can be emotionally moving and deeply motivating in ways that are of enormous importance to the speaker and hearer. The utterance can have this significance even though both speaker and hearer know that the sentence is false. Pressed in the right way—"So, you really think that love arises from cardiac activity??"—they will both back off from the claim (though probably with the annoyed observation that the questioner is missing the point). The utterance of the false sentence can have this practical significance because it's not just a free-floating falsehood uttered for the hell of it: there are *real* truths of great importance standing behind it. And, indeed, uttering this false sentence may well convey those truths *more effectively* than trying to do so using straight-talk.

This is a model that the fictionalist can usefully employ. On the assumption that no moral claims are true, we can picture a world where in certain contexts their utterance might nevertheless usefully convey information about real truths of great importance (concerning Humean values). And, indeed, uttering these false sentences may well convey those truths more effectively than trying to do so using straight-talk. Yet there need be no risk of sliding into really *believing* any moral claims—any more than one who declares "I love you with all my heart" risks sliding into believing an Aristotelian view of the physiology of emotions. Thus it seems to me that the hypothesis that we set out with—of how morality is useful as a device for bolstering one's resolve to act cooperatively—is one that remains plausible even when the attitude taken toward morality alters from belief to something more akin to make-believe.[9]

The question arises whether conservationism and fictionalism are really substantively distinct positions. (See Jaquet & Naar 2016: 204.) Both the conservationist and the fictionalist recommend that speakers declare their belief in the moral error theory when they're in a metaethical frame of mind (accepting, for instance, "It is not the case that promise-keeping is morally obligatory"), while also both recommending that the rest of the time (which is most of the time) they go along with ordinary moral judgments (e.g., "Promise-keeping is morally obligatory"). The difference is that the conservationist maintains that the latter are beliefs and assertions, while the fictionalist maintains that they are something that falls short of belief and assertion. But (one might be tempted to think) the fictionalist doesn't want the mental state in question to fall *much short* of belief, for she recommends its adoption on the grounds that it provides many of the practical benefits of genuine

---

[9] I am making here a substantive assumption that using metaphor involves make-believe—something about which there is some debate. See Joyce (forthcoming) for discussion and references. No arguments in the current chapter depend on the matter being settled one way or the other; my occasional references to "make-believe" are all dispensable.

moral belief. Indeed, the fictionalist would prefer it if the user of the moral fiction is not even aware that it is a fiction (at the time of use). This, however, raises the possibility that the fictionalist really is just discussing a state of *belief* but avoiding that label. So perhaps the disagreement between the fictionalist and the conservationist is not ultimately a metaethical one at all, but rather one about the nature of belief (and speech acts).

It is a mistake, however, to interpret the fictionalist as recommending a state that falls short of belief *but only just* (to the point that it might not fall short at all). When I claim that Jake is a "spineless snake," do I *almost believe* that he really is an invertebrate reptile? No; I come nowhere near believing this! And yet for all that I may well be virtually unaware that I'm engaged in a piece of make-believe. Consider this conversation:

A: I hear that you called Jake a spineless snake.
B: Yes, I did.
A: Do you really believe that?
B: Absolutely. He's often acted in a dishonest and cowardly fashion.
A: Mary said that she thinks that Jake *isn't* a spineless snake.
B: Well, she's wrong; she doesn't know him like I do. I've known Jake to behave in ways that prove that he's a complete snake, and I've also seen him reveal himself to be totally spineless.
A: So you believe that some reptiles are invertebrates?
B. Oh … I see what you've done there.

I'll leave it to the reader to re-run that conversation on the fictionalist's behalf, substituting the phrase "Jake acted immorally."

There is a sense in which the fictionalist can allow that the speaker believes that Jake acted immorally, just as we can ordinarily allow that the speaker believes that Jake is a spineless snake. But this sense is a kind of shorthand for something else. When we say that the speaker sincerely believes that Jake is a spineless snake, what we mean is that he or she sincerely believes that Jake is dishonest and cowardly; we are interpreting the metaphor. But there should be no temptation to conflate these senses of belief ascription; it is quite evident that they are distinct, for there is also very obviously a point of view—one quickly adopted—according to which the speaker *does not for a moment* believe that Jake is a spineless snake. (By comparison, regarding the "real" belief ascription—that the person believes that Jake is dishonest and cowardly—there is no matching point of view according to which the person doesn't for a moment believe it.) The fictionalist envisages a world where the same thing goes for moral discourse.

4. Fictionalism vs. abolitionism

The abolitionist maintains that moral discourse should go the way of talk of dragons, phlogiston, witches, and Zeus. In most of these historical cases, the abolitionist will

have faced no opposition from fictionalists or conservationists. Upon discovering that phlogiston doesn't exist, for example, the proposals that we carry on talking as if phlogiston does exist, or carry on believing that it does in everyday contexts, would have little to recommend them. The reason is pretty clear: the benefits of believing in and talking about phlogiston depend almost entirely on the assumption that in doing so we are accurately describing the way the world really is; once it is discovered that the world is not that way, we incur no significant cost and lose no significant benefit if we just drop the talk.[10] Moral discourse is different. Here the costs and benefits are considerable, and it is far from obvious that they rely on moral judgments being true, so unorthodox avenues like fictionalism and conservationism emerge as possibilities worthy of consideration.

Abolitionists could support their theory by arguing that conservationism and fictionalism would involve our incurring extra harms just in virtue of the weird kind of psychological states they recommend. The abolitionist Richard Garner imagines the error theorist trying to carry on with morality:

> Since moral judgments, we are now assuming, are false, what we say is sure to conflict with reality at many points, and then we will need to resort to evasion, obfuscation, or sophistry just to maintain our fiction. It is hard to estimate the damage this can do. If we continue to insist on the truth of our fiction, if we defend it as strongly as a convinced moralist would, then we are courting doxastic disaster, Orwellian epistemology, and perhaps a nervous breakdown. (Garner 2010: 227)

Garner's target in this passage is what he calls "assertive moral fictionalism"—but it appears that the theory he means to attack is none other than what we are here calling "conservationism" (the label of which wasn't around in 2010). If this is right, then I am fairly sympathetic with Garner's objection. Note that his concerns are quite general—presumably applying to *any* conservationist recommendation regarding not just morality but any similarly prevalent and customary (though erroneous) discourse (see Garner 1993).

It is much less clear, however, that Garner's objection, or any like it, bites against the kind of fictionalism under discussion here. Worries that fictionalism might require a kind of damaging dissonance of mental states seem predicated on the misunderstanding that fictionalism is pretty similar to conservationism—that it recommends that people come perilously close to having contradictory beliefs. But this need be no truer of fictionalists than it is true of someone who utters "Jake is a spineless snake." Does the latter person run the risk of both believing that Jake is a reptile while believing that Jake is not a reptile? Does declaring "I love you with all

---

[10] An exception might be "Zeus." There are cases to be mounted for both religious fictionalism and religious conservationism, and at some historical time and place this could have gone for Zeus and his coterie. Ovid's fictionalistesque recommendation from *The Art of Love* springs to mind: "It is expedient there should be gods, and as it is expedient let us deem that gods exist" (trans. J. H. Mozley (Loeb Classical Library, 1969: 57)).

my heart" require "evasion, obfuscation, or sophistry"? Does one court doxastic disaster or a nervous breakdown by protesting "I was *not* a pig at dinner!"? Clearly not.

One might complain that even if fictionalism doesn't involve people coming close to having contradictory beliefs, it nevertheless does encourage a damaging kind of psychological dissonance, for it recommends that a person believes that it is not the case that promise-keeping is obligatory (for example) while cultivating the emotions and motivations consistent with promise-keeping's being obligatory.

In order to see that this complaint is misplaced, remember that there is more in play in the fictionalist's psyche than just (i) the disbelief that X is obligatory and (ii) the emotions (etc.) in favor of performing X. There is also (iii) the belief that X is a good idea on non-moral (Humean) grounds. There is no dissonance here. It is (again!) like someone's disbelieving that Jake is a snake but nevertheless avoiding him. What justifies the avoidance behavior is the belief that Jake is dishonest and cowardly. To bring this analogy better into line with the fictionalist proposal, let's imagine that this person might suffer some weakness of will and feel the occasional foolish temptation enter into cooperative ventures with Jake. (Perhaps Jake is a bit of a smooth-talker.) The emotions associated with avoidance are still the appropriate ones to have, but (let's suppose) merely rehearsing dry beliefs about Jake's past bad behavior sometimes don't quite arouse those emotions. What the person finds works better is conjuring up a juicy metaphor: "Jake is such a *spineless snake*!" This metaphor has vividness and focuses attention in a way that straight-talk cannot (or cannot so well). This is, arguably, the principal general reason that metaphor has such a central place in our language and thoughts. It is not an idle foible of language or the province predominantly of poets; it is ubiquitous and constantly earns its keep.[11]

Thus I do not think that the abolitionist has much hope of discrediting fictionalism *in general terms*: showing that the type of psychological profile recommended by the fictionalist is likely to have damaging consequences. Each of us already has that type of psychological profile. Rather, the abolitionist must take on the *moral* fictionalist in particular: arguing that making a fiction of morality would be pragmatically worse than just abolishing it. And the abolitionist certainly thinks she has the resources for pressing such an argument, for she thinks that morality is on balance harmful when the object of belief, so is likely to remain harmful when the object of make-believe. Garner writes: "The bad effects of believing in moral facts are as likely to result from the pretense as from the belief" (2010: 230). Indeed, it is likely to be the very qualities touted as *benefits* by the fictionalist (and conservationist) that the abolitionist is inclined to place in the column labeled "costs." All parties can accept that one of the effects of moralized thinking is the strengthening of motivation, but where the fictionalist and conservationist thereby see a useful tool for combating weakness of

---

[11] After writing this sentence, I realized that I had unconsciously paraphrased a passage from the anthropologist Bronisław Malinowski (which years ago I quoted elsewhere) writing on the role of *myths* in human cultures. He writes of myth that "it is not an idle tale, but a hard-worked active force; it is not an intellectual explanation or an artistic imagery, but a pragmatic charter" (Malinowski 1926: 23). The connection between metaphor and myth is a very apt one to highlight.

will (inter alia), the abolitionist sees something that inflames disagreements and encourages disputants to dig in their heels and refuse to compromise.

I suspect that when all parties sit down and consider the matter calmly, they will agree on what really seems a fairly obvious truth: that morality is sometimes beneficial and sometimes harmful. And they will also, I suspect, be open to the suggestion that keeping the benefits while reducing the harms may involve neither the wholesale embrace of morality nor its wholesale elimination, but something altogether messier. In this vein, Jessica Isserow (this volume) promotes an "epistemic vigilance" based on the assumption "not only that that there are tangible benefits of moral practice, but that we can conceivably reap these benefits while controlling for the costs." Along similar lines, Björn Eriksson and Jonas Olson (this volume) endorse "negotiationism," according to which we "should normally be on the lookout for adverse effects of potential moralizing, and be prepared to retract and go for demoralization if possible." (Even Garner allows that circumstances might arise where moralizing is called for.)

I endorse these proposals that we should treat the baby and the bathwater separately if we can, but I remain more sympathetic to fictionalism remaining in the mix than do Isserow, Erikkson, or Olson. At the risk of being repetitive, let me once more suggest that we think about metaphor. Metaphors can be beneficial, but can also be harmful. Calling Jake a "spineless snake" may well usefully motivate me and others to shun him through its being more dramatic and evocative than simply referring to his personality traits and/or behavioral dispositions, but it could also be harmful: it could blind us to Jake's attempts at self-improvement, for example, or it could be self-destructive for Jake to habitually label himself in this manner, or it could be simply too vague for certain conversational purposes. There are, after all, certain familiar contexts where the use of metaphor seems misplaced and discouraged: for example, in writing up scientific experiments and in official legalistic settings. This is not to say that the very practice of metaphor is banned from these contexts—I have no doubt that a glance at real examples of scientific and legalistic language will in fact reveal a plethora of metaphor. But figurative language for key elements will be frowned upon. An anthropologist investigating cross-cultural limerence, for example, may refer to a subject's "intense emotions of intimate connection" (say), but not his "loving another with all his heart." Jake's reliability as a witness in a court case may be cast into doubt due to his track record of "dishonesty," but not his history of being "a snake."

Part of why metaphor is discouraged in these settings is, I suppose, that it evokes emotion when participants are endeavoring to remain reflective and dispassionate. Yet it is also highly plausible that it is, inter alia, its very capacity to evoke emotion that makes metaphor so valuable in other contexts. Another reason that metaphor is discouraged in these formal settings is, I suppose, that it is too vague and open to interpretation when participants are endeavoring to speak with precision of empirically tractable subjects. I would like to suggest that this is another quality that makes metaphor so valuable in other contexts: sometimes we don't *want* to speak in a

determinate way; sometimes it suits our communicative purposes much better to make vague gestures, even when talking about matters of extreme importance.

In the case of metaphor, it appears that we generally do a pretty good job of shifting back and forth between metaphor-discouraging contexts and metaphor-encouraging contexts. We do it smoothly without much internal anxiety or interpersonal confusion. This is promising news for those proposing "epistemic vigilance" or "negotiationism." But it's important to note that the example of metaphor shows that we move smoothly between embracing and dropping *a fictive attitude* toward something that we believe to be false, not that we move smoothly between embracing and dropping *a belief* toward something that we believe to be false. For example, in some contexts I will use the metaphor "Jake is a spineless snake," while in other contexts it might be more suitable for me to drop the metaphor and instead speak more carefully about Jake's personality and behavioral dispositions. We can describe this as moving in and out of a fiction, and metaphor shows it to be a commonplace practice. But it is not the case that in some contexts I will believe and assert that Jake is *literally* a spineless snake, while in other contexts I will drop the belief and assertion and instead speak more carefully about Jake's personality and behavioral dispositions.

In other words, if the abolitionist is partially right in maintaining that we should drop the moral overlay in certain contexts, but is also partially wrong because the moral overlay can usefully remain in other contexts, *and* if we look to our use of metaphor as a potential model of how something like this is already a familiar practice, then a mixture of abolitionism *and fictionalism* looks more promising than a mixture of abolitionism and conservationism.

5. Conclusion

The conclusion just reached, that there is something to be said for a mix of abolitionism and fictionalism, may seem like something of a back-down for the fictionalist—as if she has had to water down the ambitions of her original theory. But I do not believe that this is correct, for I don't think that any "un-watered-down" version of fictionalism has ever been seriously proposed—certainly not by me. To the extent that moral fictionalism is a *recommendation*, it has always presupposed the existence of non-moral values, norms, and reasons that survive the error theoretic arguments, so it has always presupposed that whatever the moral fiction recommends as morally obligatory (say) will—assuming one has chosen the right moral fiction—be strongly recommended on non-moral grounds as well. A common (but misplaced) worry about moral fictionalism is that it is difficult to see how something that one believes to be a "mere fiction" can possibly motivate serious decision-making on weighty matters. The worry is misplaced because it overlooks the serious and weighty *non-moral* considerations that are expected to lie behind the moral overlay.

The fictionalist need not claim that making a fiction of morality is *always* recommended, regardless of how a group or individual is situated, and nor need she claim that, for groups and individuals for which it *is* recommended, the moral fiction

must be maintained constantly and come what may. All that the fictionalist claims is that for typical groups of humans situated pretty much as we are, but who become moral error theorists, the practice of maintaining morality as a fiction will be, much of the time, a good idea. (And the vagueness of "much of the time" doesn't bother me in the slightest.) If it turns out that there exist serious or formal contexts where the moral fiction is best dropped and matters explicitly dealt with in literal Humean terms, that's no skin off the fictionalist's nose.

References

Campbell, E. 2014. "Breakdown of moral judgement." *Ethics* 124: 447-480.

Cuneo, T. & Christy, S. 2011. "The myth of moral fictionalism." In M. Brady (ed.), *New Waves in Metaethics*. NY: Palgrave Macmillan. 85-102.

Garner, R. 1993. "Are convenient fictions harmful to your health?" *Philosophy East and West* 43: 87-106.

Garner, R. 2010. "Abolishing morality." In R. Joyce and S. Kirchin (eds.), *A World Without Values: Essays on John Mackie's Moral Error Theory*. Dordrecht: Springer. 217-233.

Ingram, S. 2015. "After moral error theory, after moral realism." *Southern Journal of Philosophy* 53: 227-248.

Jaquet, F. & Naar, H. 2016. "Moral beliefs for the error theorist?" *Ethical Theory and Moral Practice* 19: 193-207.

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

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

Joyce, R. 2017. "Fictionalism in metaethics." In D. Plunkett & T. McPherson (eds.), *Routledge Handbook of Metaethics*. NY: Routledge. 72-86.

Joyce, R. Forthcoming. "Fictionalism: Morality and metaphor." In F. Kroon & B. Armour-Garb (eds.), *Philosophical Fictionalism*. Oxford: Oxford University Press.

Kalderon, M. 2005. *Moral Fictionalism*. Oxford: Oxford University Press.

Lutz, M. 2014. "The 'now what' problem for error theory." *Philosophical Studies* 171: 351-371.

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

Malinowski, B. 1926. *Myth in Primitive Psychology*. London: Kegan Paul.

Olson, J. 2011. "Getting real about moral fictionalism." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Vol. 6*. NY: Oxford University Press. 182-204.

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