

***Moral fictionalism: When falsehoods are too useful to throw out***  
**Richard Joyce**

Penultimate draft of article appearing in *Philosophy Now* 82 (2011)

The history of moral philosophy can seem a disappointing spectacle. Large tracts of the project can be interpreted as thinkers taking their own moral preferences and trying desperately to prove them *correct*. The most ambitious among them will aspire to *objective absolute necessary truth* for their moral claims. Failing this (as fail it does), various softer projects will be countenanced: Perhaps *objectivity* is asking too much; maybe we can allow that moral truths are in some manner constituted by human practices. Or perhaps *absolutism* can be dropped; maybe we can be satisfied with one moral truth *for us* and another moral truth *for them*. And so on. The theories are plentiful, the convolutions byzantine, the in-fighting bitter, the spilt ink copious, and the progress astoundingly unimpressive.

It is a disappointing spectacle not merely because we have so little to show for it after two and a half thousand years, but because this enterprise of self-exoneration can seem immature—as deriving from an anxious need for reassurance. Yet when one reflects on where our moral judgments come from, it appears unlikely that this reassurance will be forthcoming. By the time we take our first tottering steps, each of us is already immersed in a social world rich in concepts like *right* and *wrong*, *desert* (relating to rewards and punishments), *must* and *mustn't*. Our childhood is one grand advertising campaign designed to get us to internalize these concepts and take them seriously—a campaign, moreover, that in all likelihood we are biologically designed to find compelling (because thinking in this fashion helped our ancestors produce more babies than their competitors). And so we do.

Nowhere, however, does this account of how we come to make moral judgments presuppose that any of the beliefs in question are actually *true* (even approximately so); and now, as adult philosophers—being in a position to stand back and see the process for what it is—do we really need to concoct cunning theories designed to earn this missing truth for our moral beliefs? Even supposing this mission were possible (regarding which the fat back catalog of failed theories cautions pessimism), the motivation to demonstrate that one's current moral framework—the product of a particular contingent cultural setting interacting with a particular contingent evolutionary trajectory—is more-or-less *correct* (even if only “correct *for us*”) seems an expression of grotesque hubris. Even if this felt need for reassurance is natural, investigating why we feel the need seems the braver course than seeking its satisfaction.

But what, one might protest, is the alternative? Surely it is not enough simply to make judgments about what is morally right and wrong (virtuous, evil, etc.) and think no further on the matter; surely these judgments require *justification*. How can I in good faith make the judgment that something is morally wrong if I cannot also maintain that this judgment is *true*? And if I'm going to maintain that it's true then I shall require confidence that there exists some sensible account of those facts in virtue of which it is true.

This is not just some abstruse philosophical quandary; this is serious! There is a very real possibility that no reasonable account can be given of what it takes for a moral judgment to be true. There is, thus, a very real possibility that none of our moral judgments are true at all: It is untrue that punching babies is morally wrong; it is untrue that keeping promises is morally

better than breaking them; it is untrue that we have any moral duties towards our fellow humans whatsoever. Someone who maintains this skeptical position is called a “moral error theorist.”

The error theorist holds that our moral judgments purport to be true but that the universe just doesn’t contain the requisite properties (goodness, wrongness, etc.) necessary to render any of them true. The moral error theorist can be thought of as taking an attitude towards morality analogous to the attitude an atheist takes towards religion. This kind of moral skeptic typically doesn’t attribute error to *all* evaluative talk, but just to *moral* talk. Many moral terms have non-moral uses, and with the latter the error theorist has no beef. Thus, he still can talk without flinching of “the *best* computer,” “the *wrong* direction,” “a *good* idea”; he can maintain that someone “*should* not have another glass of wine,” “*ought* to believe X (given the evidence),” “has legal *obligations*,” and so forth. (Analogously, just because the atheist doesn’t believe in God’s commands, it doesn’t follow that she doesn’t believe in Fred’s commands.) But as such linguistic practices shade into their *moral* usages, the error theorist protests. The aim of this article is not to elucidate or defend such a position, but to wonder what might happen next.

If the error theorist is correct, then the natural assumption is that we should eliminate moral considerations from our minds entirely, just as we have eliminated beliefs in mermaids, phlogiston, and astrology. But that, surely, is an appalling possibility—a prospect so dreadful that any amount of labor to avoid it is hardly an immature yearning for comfort, but rather a heroic stand against forces that threaten to unhinge civilization.

Yet would it really be so awful to do away with morality? I don’t think so. If there *is* a scary idea hereabouts, it’s the assumption that all that stands between our present gentle existence and a bedlam of savagery are our moral beliefs. Thankfully, that assumption is mistaken; rather, our reasons for engaging in the myriad of cooperative ventures typically reckoned to be supported by morality are robustly overdetermined. Even if I cease to think of punching babies as morally wrong, it’s not as if I’m suddenly filled with the desire to do so; it’s not as if I suddenly lack any reason to refrain from doing so; it’s not as if I’ll stand idly by if I see someone else doing it. Contrary to a popular but almost entirely unexamined image, the moral error theorist may be as much a friendly, trustworthy, upstanding citizen as anyone else. True, he doesn’t believe that punching babies is morally wrong; but then again nor does he believe that it’s morally right or morally permissible either; he doesn’t believe it’s *morally* anything.

Moral thinking, we can assume, evolved because it served some range of purposes. A very coarse-grained but plausible hypothesis is that this kind of thinking in some manner strengthened ancestral social cohesion in a reproductively beneficial way. (Whether this “reproduction” is the making of babies or the persistence and proliferation of cultural practices is something about which I will remain neutral. Also, I am using “social cohesion” as an extremely broad umbrella term.) But do we now *need* moral thinking for social cohesion? Just as one can make an autonomous decision not to maximize one’s genetic reproductive potential—after all, not many men are queuing at the sperm bank every day eager to max out their donations—so too we can make an autonomous decision to support social cohesion. (If we already take an interest in it, we can make a decision to endorse that interest.) We can simply decide to be kind to each other, to refrain from harming each other, to repay our debts, and so on. If we appreciate social cohesion and the benefits it brings, then why do we need morality prodding us along to do these things? “But what about people who

*aren't* interested in social cohesion?" I can imagine someone complaining. And my answer is that of course there will be individuals who don't choose in this manner, but when those things that support social cohesion get the backing of morality there are individuals who don't choose them then, too. The amoral society is not forced to tolerate socially destructive behavior any more than the moral society.

So I don't have much of a problem imagining humans getting on very well in a moral-judgment-free world. It is, in fact, not far-fetched to expect that in certain respects they will do even better than us. (No more moral concepts to be manipulated in the service of warfare, for example.) But it is also reasonable to suspect that in certain respects they might do worse than us. After all, moral thinking is a well-honed tool that does an important job; it suits our psychological configuration; we are familiar with its contours. Eliminating it entirely from our thoughts and language can be expected to have some kind of deleterious impact on our motivations.

Epistemological principles (concerning what one ought and ought not believe, given one's other beliefs) require any committed moral error theorist to discard his moral beliefs. But a group of error theorists sensitive to these principles may then consider themselves in a bit of a pickle, for at the same time *practical* considerations are telling them to avoid the costs that will be incurred by giving up these beliefs. *Moral fictionalism* is a view designed to help them out of this pickle. Moral fictionalism maps a stance according to which these error theorists can respect all the decrees of epistemology while offsetting some of the costs involved in eliminating moral discourse from their lives.

Suppose Mary was, as a teenager, foolish enough to believe in astrology. She believed, among other things, that sagittarians are prone to optimism and self-confidence. But (thankfully) Mary has grown up and realized how preposterous are the causal presuppositions upon which astrology is based, and how worthless is the evidence touted in its favor (e.g., all those optimistic sagittarians). One can expect that she will cease to have astrological beliefs—she will, for example, cease to believe that sagittarians are characteristically optimistic. She is, moreover, epistemologically required to cease to believe this, inasmuch as we are supposing that she now steadfastly believes that it is false, and thus for her also to believe it would be contradictory.

But must Mary *never again utter the sentence* "Sagittarians are characteristically optimistic"? To answer in the positive seems hasty. One can imagine various scenarios where Mary should still utter the sentence. If someone holds a gun to her head and commands her to say it, for example. Or maybe she's just joking with another astrology-skeptic and she offers this as a tongue-in-cheek explanation of a colleague's actions. Or perhaps she's surrounded by ardent astrologers and has a sound reason to refrain from insulting them. (Of course, in most such situations one would still prefer to hear her say "I just don't believe in this nonsense," but one can easily flesh out the example in such a way that this preference dissolves and we are more inclined to sympathize with Mary's decision to play along.) We can imagine a whole array of circumstances where Mary utters this sentence—the content of which she doesn't believe—and we would not fault her for it. In some of these circumstances she will be simply uttering the sentence without asserting it (e.g., when joking), while in others she will be asserting it (in particular, where she believes that her audience will assume that she believes what she says). In the latter scenarios we may admit that Mary is *lying* (since she asserts something she doesn't believe)—but even so we may often see that it's a harmless white lie, and thus won't condemn Mary for any great misconduct. Of course, it doesn't take much

effort also to imagine circumstances wherein Mary utters the sentence and we are inclined to criticize her for dishonesty, bad faith, or hypocrisy. My point is that there is a great variety of cases where Mary carries on occasionally using astrology talk, and it would be a mistake to tar them all with the same condemnatory brush.

Now let's return to the moral error theorist—call him "David." David doesn't believe that punching babies is morally wrong, but we can imagine various situations in which he'll have good reason to utter the sentence "Punching babies is morally wrong." Imagine that David is surrounded by a population who do believe in moral wrongness and believe that baby-punching has it. We should remind ourselves that David is no fan of baby-punching; in fact, the thought of it sickens him. He believes that baby-punching ought to be prevented and perpetrators severely dealt with. He thinks all this on *non-moral* grounds. So when a moral believer asks him his opinion of baby-punching, David *could* embark on a long and likely-to-be-horribly-misunderstood explanation of his non-moral grounds for opposing the action. But would we really accuse him of any great transgression if he simply says "Baby-punching?! Oh, that's just morally *wrong!*"?

I, for one, sympathize with David's decision. Even acknowledging that he has lied doesn't lessen my sympathy, for I can see (from my God's eye perspective) that it's an understandable and harmless lie. David didn't really mean to mislead his audience; he judged that the "spirit of the conversation" called for him to say something he doesn't believe rather than embark on an unorthodox exposition that might ultimately involve him having to explain esoteric concepts like the difference between hypothetical imperatives and categorical imperatives, or the distinction between motivating reasons and normative reasons. (Don't worry if you're not familiar with these terms; that's the point. Most people don't know the difference, and so it's expedient for David to bypass the lengthy explanation.) Thus for the sake of brevity, for the sake of engaging with the spirit of the conversation, for the sake of giving the questioner the least confusing and misleading answer, David tells a lie. Big deal.

By contrast, if the questioner were really asking about David's *philosophical* views—if he were interested in whether David is an error theorist or a moral realist, and chose to pose this broader question using baby-punching as an example—then that would be a different matter. In that conversational context, if David answered "Punching babies is morally wrong" then he would be misleading the questioner in a more serious way, and I would be a lot less inclined to say "Big deal." (Of course, David, being an error theorist, doesn't believe that there is anything *morally* wrong with lying even in serious and harmful ways; but we will assume that he is opposed to it on non-moral grounds, just as he is opposed to baby-punching.)

What I have just described is a mundane defense of a fairly mundane form of moral fictionalism. It grants that a moral error theorist may continue to say things that he doesn't believe (on occasion even assert them) on grounds of an unremarkable kind of expediency, while not being the subject of accusations of bad faith (etc.)—no accusations, in any case, over which he need lose any sleep. Moral fictionalism gets more interesting and more controversial when it recommends that the error theorist "immerse" himself in the fiction in a rather more full-bodied fashion, such that he doesn't just carry on using moral language, but also "accepts" these falsehoods in a way that allows them some positive active influence upon his decisions and motivations. This needn't be good advice for *all* the people *all* the time; it is sufficiently interesting if it promises to produce practical benefits for some of the people some of the time.

An important feature of fictive immersion is that when immersed one is distracted from the fact that one is an error theorist. Hence much of the time one speaks and acts and even thinks as if one really believes it all. The idea is, then, that one can in this way gain some of the pragmatic benefits that come from sincere moral belief. And yet in doing so one violates no epistemological imperative since one *doesn't* believe it; one's "acceptance" of morality falls short of belief since one remains disposed to concede, if pressed in an appropriately serious and critical way, that it's all false. (By analogy, I violate no epistemological dictum if I happen to daydream something that I know to be false.) We can call this fictive attitude a kind of "pretense," with the qualification that it may be an important and highly-played kind of pretense. But how can mere pretense have a sufficient impact on one's motivations so as to affect significant costs and benefits? The quick answer to this question (as to so many others) is: Humans are strange creatures.

One of the oddest things about humans is our interest in fictions. It is possible to imagine intelligent creatures who feel no emotion when engaging with characters who never existed and narratives that never happened. But that's not us. We are creatures who, for whatever reason, can be moved by fictions. Stories and images affect our moods, emotions, and motivations. We are also creatures whose moral judgments can be affected by the smells in our environment, or whether we're in a messy room, or in a dark room, or whether we've recently washed our hands. A multi-billion dollar advertising industry is all too keenly aware of how a catchy jingle or ambient muzak can influence us to part with hard-earned cash.

The error theorist who immerses himself in a moral fiction takes advantage of these quirky aspects of his own human psychology. He cultivates a habit of bringing moral concepts to bear on practical problems; he allows moral emotions like disgust, anger, and guilt to wash through him; he is acutely familiar with classifying his social environment in moral terms. All going according to plan, this moral fiction doesn't encourage him to do anything that he wouldn't upon reflection choose to do anyhow on non-moral grounds. But moral thinking has some advantages over careful non-moral thinking, in that it can be fast and frugal, less prone to self-sabotaging rationalizations, able to banish practical calculation from the decision procedure when the very act of calculating is suboptimal.

Earlier we imagined David the lone error theorist surrounded by moral believers. Now let's imagine a whole population of moral error theorists, who nevertheless habitually immerse themselves in the moral fiction for the various practical benefits that doing so brings. Nobody is claiming that in taking this stance they enjoy *all* the advantages of sincere moral belief. But the stance pays for itself so long as it delivers *some* of those benefits (with no extra costs). An interesting result is that so long as all members of the population are, at some level, "in on the game," then they have in effect invented a linguistic convention according to which when someone says, e.g., "Punching babies is morally wrong," nobody takes this as an expression of a real belief, but rather as a kind of expedient shorthand for something non-moral (but no less important for that). Perhaps it is best interpreted as a kind of publicly sanctioned pretense—a pretense maintained by speaker and audience—in which case it will not be functioning as an *assertion*, which is to say that it will not *be* an assertion (and thus not a lie). In such a case, the moral fictionalist is an error theorist who neither believes nor asserts moral propositions, but nevertheless allows morality an active role in his life.

And what would be so terrible about such a linguistic convention? Should it be subject to intellectual censure any more than, say, our present conventions surrounding metaphor, sarcasm, euphemism, litotes, or synecdoche? If, furthermore, it serves some practical function

(which is an empirical matter that has been only gestured towards here) then it cannot be obviously faulted on pragmatic grounds. Willard Van Orman Quine may once have spoken with withering contempt of “philosophical double talk which would repudiate an ontology while simultaneously enjoying its benefits” (*Word and Object*, 1960, p.242), but, when it comes right down to it, what does this accusation of “double talk” amount to beyond highfalutin name-calling?