*Moral fictionalism**
Richard Joyce

Were I not afraid of appearing too philosophical, I should remind my reader of that famous doctrine, supposed to be fully proved in modern times, "That tastes and colours, and all other sensible qualities, lie not in the bodies, but merely in the senses." The case is the same with beauty and deformity, virtue and vice. This doctrine, however, takes off no more from the reality of the latter qualities, than from that of the former; nor need it give any umbrage either to critics or moralists. Though colours were allowed to lie only in the eye, would dyers or painters ever be less regarded or esteemed? There is a sufficient uniformity in the senses and feelings of mankind, to make all these qualities the objects of art and reasoning, and to have the greatest influence on life and manners. And as it is certain, that the discovery above-mentioned in natural philosophy, makes no alteration on action and conduct; why should a like discovery in moral philosophy make any alteration?

David Hume, "The Sceptic" in *Essays, Moral, Political, and Literary* (1742)

1. IF THERE'S NOTHING THAT WE MORALLY OUGHT TO DO, THEN WHAT OUGHT WE TO DO?

On the very last page of his book *Ethics: Inventing Right and Wrong*, John Mackie (1977) suggests that moral discourse—which he has argued is deeply error-laden—can continue with the status of a "useful fiction." I presume that most people will agree, for a variety of reasons, that morality is in some manner useful. The problem, though, is that its usefulness may depend upon its being *believed*, but if we have read the earlier stages of Mackie's book and have been convinced by his arguments, then surely the possibility of believing in morality is no longer an option. Even if we somehow *could* carry on believing in it, surely we should not, for any recommendation in favor of having false beliefs while, at some level, knowing that they are false, is unlikely to be good advice. So how useful can morality be if we don't believe any of it?

This paper will assume without discussion that Mackie's arguments for a moral error theory are cogent (or, at least, that their conclusion is true). This amounts to assuming two things: first, that moral discourse typically is assertoric (that is, moral judgments express belief states); second, that moral assertions typically are untrue. Mackie's particular argument holds that the problems of morality revolve around its commitment to Kantian categorical imperatives: morality requires that there are actions that persons ought to perform regardless of their ends. But, Mackie argues, such imperatives are indefensible, and therefore morality is flawed. A moral error theorist must hold that the problematic element of morality (categorical imperatives, in Mackie's opinion) is *central* to the discourse, such that any "tidied up" discourse, one with the defective elements extirpated, simply wouldn't count as a *moral* system at all.

There are rich and inventive arguments against Mackie, but here we will suppose them all to fail. The question that this paper addresses is "What, then, ought we to do?" Mackie's answer appears to be "Carry on with morality as a fiction," and it is this possibility that I wish to examine closely. The aim is to understand what such an answer

may mean, and to attempt a defense of it. I will call the view to be defended "moral fictionalism." Fictionalism promises to be a way by which we can avoid the situation that Quine so deplored, of employing "philosophical double talk which would repudiate an ontology while simultaneously enjoying its benefits" (1960: 242). Note that fictionalism is not being suggested as something that is true of our actual moral discourse; rather, it is presented as a stance that we could take towards a subject matter—morality, in this case—if we have become convinced that the subject is hopelessly flawed in some respect, such that we cannot in good conscience carry on as before. In the useful terminology of John Burgess, I am peddling a "revolutionary" not a "hermeneutic" fictionalism (1983[1]).

One might think that the question "If a moral error theory is the case, what should we do?" is self-undermining. And so it would be, if it were asking what we *morally* ought to do, but that is not what is being asked. It is just a straightforward, common-or-garden, *practical* "ought." The answer that the question invites will be a hypothetical imperative, and we will assume that whatever arguments have led us to a moral error theory have not threatened hypothetical imperatives. (In other words, to hold a moral error theory is not to hold an error theory for practical normativity in general.) I do not want this issue to depend on any particular view of how we make such practical decisions. Let us just say that when morality is removed from the picture, what is practically called for is a matter of a cost-benefit analysis, where the costs and benefits can be understood liberally as preference satisfactions. By asking what *we* ought to do I am asking how a *group* of persons, who share a variety of broad interests, projects, ends—and who have come to the realization that morality is a bankrupt theory—might best carry on. (Two comments: (1) I wouldn't object if we decided to speak of *informed* rather than actual preferences; (2) no assumption is being made that preferences will be selfish in content.)

I will begin by discussing fictionalism in general, outlining how it might be that a person might carry on using a discourse that she has come to see as flawed. It will be useful if initially we avoid the distractions that the particular case of *moral* fictionalism might bring, and so I will begin by discussing an example that in some ways is less controversial: color fictionalism.

2. CRITICAL CONTEXTS

Suppose that after reading some eighteenth-century philosophers David comes to endorse an error theory about color. We needn't go into the arguments that might lead him to this conclusion, but they probably have something to do with the thought that one of the central platitudes about color is that it is a type of surface property of objects with which humans can have direct acquaintance (e.g., with normal eyesight on a sunny day), coupled with the thought that there simply aren't any properties like *that*. In other words,

---

[1] Burgess' original distinction was between two forms of nominalism, but was extended to fictionalism in his (1997), with Gideon Rosen. For criticisms of hermeneutic fictionalism, see Stanley (2001).

for philosophical reasons he ceases to believe that the world is colored in the way that it appears to be colored, which (further philosophical reasons lead him to think) implies that it is not colored at all. Maybe he is confused in coming to such a conclusion, but that is not the issue.

The issue is: given that he has come to have this philosophical belief (however confusedly) what happens to all his color discourse? Does he stop saying things like "The grass is green"? If someone asks him what color his mother's eyes are, does he reply that they are no color at all? Does he cease to appreciate sunsets or Impressionist paintings? Does he wear clashing clothes (while denying that anything really clashes with anything)? Of course not. In 99% of his life he carries on the same as everyone else. His vision is the same, his utterances about the world are the same, and even what he is thinking while making these utterances is the same. It is only in the philosophy classroom—moreover, only when discussing sensory perception—that when pressed on the question of whether the grass is green David might look uncomfortable, squirm, and say "Well, it's not *really* green—nothing is *really* green." This may seem like an uneasy position for him to be in. Sometimes—99% of the time, let's say—he is willing to utter "The grass is green," "The sky is blue," etc., while at other times—1% of the time—he is inclined to deny these very same propositions. Which does he believe?

It seems to me that in this case what he affirms 1% of the time determines his beliefs. Why? Because the circumstance in which he denies that the world is colored—the philosophy classroom—is the context within which he is at his most undistracted, reflective and critical. When one thinks critically one subjects one's attitudes to careful scrutiny ("Is my acceptance of *p* really justified?"); robust forms of skepticism are given serious consideration; one looks for connections and incoherencies amongst one's attitudes; one forms higher order attitudes towards one's first order judgments. It is important to see that this distinction between more critical and less critical contexts is asymmetric. It's not merely that a person attends to *different* beliefs when doing philosophy than when, say, shopping; nor that she questions everyday thinking when doing philosophy, but equally questions philosophy when shopping. Critical thinking investigates and challenges the presuppositions of ordinary thinking in a way that ordinary thinking does not investigate and challenge the presuppositions of critical thinking. Critical thinking is characterized by a tendency to ask oneself questions like "Am I really justified in accepting that things like shops exist?"—whereas the frame of mind one is in when shopping is *not* characterized by asking "Am I justified in accepting that there is some doubt as to whether shops exist?"

This notion of what a person is disposed to assent to if placed in a critical context must not be read as involving any far-fetched counterfactual idealization. Who can judge what manner of bizarre things one would assent to if given *perfect* powers of reflection and critical thinking? A person's "most critical context" must be fixed in actuality—and the obvious means of achieving this grounding is to stipulate that he must sometimes (at a minimum, at least once) have *actually inhabited* that context, and therein either assented to, or dissented from, the thesis in question. In other words, it would be too bizarre to hold that an individual, who has never given the issue any careful thought whatsoever,

but thinks and acts in accordance with theory T, does not really believe T simply because if he *were* to think carefully about it, he would deny it. But if we add that at some point he *has* adopted a critical perspective and therein sincerely denied T, and remains disposed to deny T were he again to adopt that perspective, then he disbelieves T, regardless of how he may think, act, and speak in less critical perspectives. In David's case, his most critical context is philosophical thought—thus, though he occupies this position only 1% of the time we're supposing, it is his pronouncements therein that reveal his beliefs. The rest of the time he still *has* this skeptical belief, but he is not attending to it. Nevertheless, *all* the time David remains disposed to deny that the world is colored if placed in his most undistracted, reflective and critical context, thus *all* the time this is what he believes.

3. FICTIVE JUDGMENTS

This leaves us with the question of how we should describe David's color claims in that 99% of his life where he utters propositions (e.g., "The grass is green") that he disbelieves. We can begin by reminding ourselves of a more familiar circumstance in which people utter propositions that they disbelieve: story-telling. When I utter the sentence "There was once a goblin who liked jam" as part of telling a story, I am not expressing something that I really believe. If pressed in the appropriately serious way ("You don't *really* believe that there was once a goblin who liked jam, do you?") then I will "step out" of the fiction and deny those very propositions that a moment ago I was apparently affirming.

Some people have argued that sentences concerning fiction ought to be interpreted as containing a tacit story operator, such that they maybe treated as true assertions; thus the sentence "There was once a goblin who liked jam" may be used to express the true proposition "According to Hans Christian Andersen's story, there was once a goblin who liked jam." (See, for example, Lewis, 1978.) This is inadequate as a general claim, for it fails to distinguish two different things that we can do with a story: describing the story versus telling the story. When we tell a story we are pretending something: that we are a person who has access to a realm of facts that we are reporting. (We might also partially pretend to be characters in the story, which is why we will speak their parts in a gruff or squeaky voice.) But if every sentence of the story uttered contained an unpronounced fiction operator, then there is no sense to be made of the claim that the storyteller is pretending. (How would one *pretend* that according to Hans Christian Andersen's story, there was once a goblin who liked jam?[2]) This is not to deny that on occasions the proposition "According to Hans Christian Andersen's story, there was once a goblin who liked jam" might be expressed elliptically, minus the prefix, but this is not what we are doing when we *tell* the story. On such occasions we are not asserting anything, but *pretending* to assert.

---

[2] Walton (1978) makes a similar point.

The same distinction can be made regarding skeptical David's color claims. When, in ordinary conversation, he utters the sentence "The grass is green," we could interpret this as a kind of shorthand way of asserting something like "According to the fiction of a colored world, the grass is green" *or* we could interpreting him as not asserting anything at all, but rather doing something rather like engaging in a make-believe: pretending to assert that the grass is green. I prefer the latter interpretation. It is true that at the moment of making the utterance it doesn't *seem* to David as if he is participating in an act of pretence, but nor does it seem to him as if he's making an implicit reference to the content of a well-known fiction. The matter won't be settled by asking David what he takes himself to be doing. Unless we force him into the philosophical context where he denies the existence of colors altogether, then asking him in an ordinary context whether he is asserting that the grass is green is likely to meet with an affirmative answer. But *that* claim—"Yes, I am asserting that the grass is green"—may be just another part of the fiction. (A Roald Dahl story, recounting many fantastic events, contains an explicit declaration that the story is not a fiction, but it's *all true*. The declaration of truth is no less part of the make-believe than the rest of the story.[3]) The issue of whether David's everyday utterance "The grass is green" is an assertion about a fiction or a fictional assertion is not an issue about how things feel to him—it is to be settled by philosophers providing an interpretation that construes David's linguistic practices most charitably.

The former interpretation—the "tacit story operator view"—does him no favors. One problem is that it cannot account for the fact that when in a more critical context David will explicitly *overturn* what he earlier claimed—he might say "What I said earlier was, strictly speaking, false." But if what he said earlier concerned the content of the fiction of a colored world, then he does not think it was false at all. A second problem with this interpretation is that it fails to make sense of the ways David might employ a color claim in a logically complex context (see Vision, 1994.) For example, he might endorse the following argument:

P1   Fresh grass is green
P2   My lawn is made of fresh grass
C   Therefore, my lawn is green

But if the first premise is elliptical for "According to the fiction of a colored world, fresh grass is green," then the argument is not valid at all. There is room for maintaining that the argument would be valid if all three claims were so prefixed, but the problem then would be that the revised second premise ("According to the fiction of a colored world, my lawn is made of fresh grass") seems so obviously false that it is surely not what David asserts when he utters P2. The fiction of a colored world, in so far as it has a determinate content at all, does not include claims about what anybody's lawn is made of (see

---

[3] Dahl's story is "The Wonderful Story of Henry Sugar," in case you're interested. Balzac's *Le Père Gorio*t also famously claims of itself that it is neither a fiction nor a romance, but "ALL IS TRUE."

comments by Lewis, 1978: 38-9).

To this it might be objected that the operator is being interpreted incorrectly. If "according to..." means not "it is claimed by..." but something more like "it is true in the fiction of...," then perhaps we might after all allow that according to the fiction of a colored world my lawn is made of fresh grass. In much the same way we might allow (indeed, insist) that it is true in the fiction of the Conan Doyle stories that humans do not have long hairy tails, that 6+5=11, that Ireland is to the west of Britain, and so on, despite the fact that one will not find such things *claimed* by the stories (nor even—with, perhaps, the exception of the arithmetical truth—*implied* by anything claimed by the stories).

But this objection leads to unsightly consequences. Suppose David just casually asserts "My lawn is made of fresh grass." Since this assertion may at any time be pressed into service as the premise of an argument (the other premises of which include color claims), if the resulting argument is to be valid we will have to interpret him as *really* having asserted "It is true in the fiction of a colored world that my lawn is made of fresh grass." But the very same assertion may be employed by David as a premise in another argument that involves no color claims and no obvious fictionalizing: he may combine it with "Fresh grass is a type of vegetation," for example, to get the conclusion "My lawn is made up of a type of vegetation." In order for this new argument to be valid we had better interpret this new premise (and the new conclusion) as also bearing the prefix. In fact, any assertion that David makes might be combined with color claims as a premise of an apparently valid argument, and so if we're to maintain that apparent validity is real validity, we're going to have to interpret everything that he asserts about anything as having this unpronounced prefix. Things get worse still if we remind ourselves that color may not be the only fiction that David participates in. Eighteenth-century philosophy may also lead him to endorse an error theory for sound and smell, for causation, for virtue and vice, and thus in order for all his apparently unremarkable, apparently valid argumentative moves to be genuinely valid, we will have to interpret every claim issuing from his mouth as brimming with unspoken prefixes.

All such unpleasantness is avoided if we do away with tacit operators, and simply interpret David's utterance "Fresh grass is green" as a kind of make-believe assertion. The content of the proposition doesn't change, any more than when I say (as part of telling a story) "There was once a goblin who liked jam" I am using "jam" with some special meaning. The sentence "There was once a goblin who liked jam" has exactly the same content whether it is used as part of a fairy tale or to foolishly assert something false. What changes is the "force" with which it is uttered. When asserting it I am presenting it as something that I believe, and putting it forward as something that my audience should believe. Linguistic conventions decree that when it has been preceded by "Once upon a time...," all such expectations are lifted.

What are we to make of an argument when some of the premises are uttered as an act of make-believe (e.g., as lines in a play) while others are straightforward assertions? Since the presence or absence of assertoric force doesn't affect the content of the premises, then if the argument was valid with its components asserted, it will be valid

with them unasserted, and remain valid if some of the components are asserted and some of them are not. For example, the following is a valid argument:

P1   It is cold tonight
P2   It is the height of summer
P3   A cold night in the height of summer is unusual weather
C   Tonight is unusual weather

If a logic teacher recited this argument to a group of incoming undergraduates as an example of validity, she would not be asserting any of the premises or the conclusion—but it would be no less valid for that. Alternatively, suppose that P1 is the line of a play, and the actor duly utters it while on stage, during a performance given on a hot summer's night. After the play, when pressed on climatic issues (curiously), he assents in all seriousness to P2 and P3. Clearly this person has not committed himself to the conclusion (which he may believe to be completely false), for the reason that he did not commit himself to P1. On the other hand, there is nothing to prevent him from "going along" with the pretence if for some reason he wants to, combining P2 and P3 with the make-believe P1, and endorsing the conclusion as part of a fictional act. If he does so, there will be no need to reinterpret his attitude to P2 and P3. These were asserted, and in asserting them he has committed himself to certain other conclusions (e.g., "If it were cold tonight, that would be unusual weather"), and may combine them with further asserted premises to yet further conclusions. In other words, unlike with the tacit operator account we do not have to interpret David's ordinary claim "My lawn is made of fresh grass" as anything other than it appears to be, let alone extravagantly reinterpreting all his other ordinary assertions that are not color claims.

Let us say, then, that David is not only an error theorist about color, but also a fictionalist. He does not believe in color, but he continues to employ color discourse. His color claims are fictive judgments, which we may think of as a kind of "make-believe"—though one should be wary of the term, since the paradigm examples that it tends to bring to mind are of rather trivial activities (pretending that the puppet is talking, make-believing that the sofa is a boat, etc.). But there is no obvious reason to assume that make-believe is always a trivial business;[4] indeed, an important objective of this paper is to convince you otherwise. We have not specified David's reasons for making these fictive color judgments—let us just say that he finds it convenient to do so. This practical value need be nothing more than the convenience of carrying on in the manner to which he has grown accustomed.

Since David is capable of overturning his everyday color discourse whenever he enters

---

[4] Autistic children fail to participate properly in games of make-believe, and this corresponds to, and arguably contributes to, a whole range of serious disabilities. See Baron-Cohen (1987); Jarrold et al. (1996). For discussion of the evolutionary importance of make-believe play in humans, see Steen and Owen (2001).

a more critical frame of mind, we should hardly describe him as suffering from self-deception. He is no more self-deceived than is someone caught up in a good novel. I suppose that the term "self-deception" *could* be applied to an ordinary person engaged in a novel, but (A) it would be an uncomfortable stretch, and (B) it would merely show that self-deception need not be in the least pernicious.[5] It is much better, I think, to distinguish being "caught up" in a fiction from being "deceived" by a fiction. A person deceived by a fiction is someone who might walk up and down Baker Street wondering where Holmes lived, or who tries to research Madame Bovary's ancestry, or who rushes on to the stage to save the princess. Fans of Sherlock Holmes do travel to Baker Street, of course, and they may well picture their hero there in the nineteenth century, but they know very well (most of them, I hope) what they're doing. At any time, if asked in all seriousness whether Holmes walked these streets, they will answer "No." They are not deceived and therefore not self-deceived; they are merely caught up in a fiction. It is the person who is incapable of dropping the fiction, who continues to speak of Holmes as a historical character even when in her most critical context, who is deceived (though further criteria would need to be met before we would describe such a person as *self*-deceived).

4. NONCOGNITIVISM AND THE LONE FICTIONALIST

If by "noncognitivism" we mean the view that a certain discourse does not typically consist of assertions, despite normally coming in the indicative mood, then it would appear that we ought to be noncognitivists about David's fictive color claims. Remember that fictionalism is being considered here as something that we could *do* with a problematic discourse, not as an analysis of any actual discourse (problematic or otherwise), thus the same goes for the consequent noncognitive stance: it is a description of a discourse that we might choose to adopt, not a description of an actual discourse. Another thing to note is that although over the years we have grown used to the idea of noncognitivists offering a "translation" of allegedly problematic everyday sentences into some unproblematic idiom, that is not what is being suggested here. For example, we are familiar with moral noncognitivists telling us that a claim like "Stealing is wrong" really amounts to "Stealing: boo!" or "I disapprove of stealing; do so as well!" One might misread the present noncognitivist proposal as suggesting in the same spirit that someone who claims "Stealing is wrong" is really saying something like "Let's pretend that stealing is wrong"—thus making it clear that the claim is not really an assertion. But this would be, as I say, a misreading. When playing a game of make-believe with children—say, crawling around on the floor pretending to be a bear—one might say, in a gruff voice, "I am a bear; I am going to eat you!" It would be an odd theory that identified the

---

[5] "Self-deception" is a contested term. In this paper I avoid any theoretical commitment on the issue, though I should say that on other occasions I would object to the term being stretched to the extent considered.

true content of this utterance as "Let's pretend that I am a bear; let's pretend that I am going to eat you." Someone saying such things would hardly be "playing a game" at all. He might as well start out saying (in an ordinary voice) "Let's pretend that I am speaking in a gruff voice." With noncognitivism defined as above, it is not incumbent on its proponents to provide a translation scheme from problematic language to unproblematic. For the moral fictionalist/noncognitivist, the content of "Stealing is wrong" is exactly what it appears to be—with whatever erroneous implications she thinks that it has remaining in place. What is different about her utterance of the sentence is the force with which she utters it.

There is, however, a troubling consequence of this kind of noncognitivist proposal, for notice that I claimed that we should be noncognitivists about *David's* fictive color discourse, implying that we might not be noncognitivists about everyone else's color claims. Noncognitivism, thus, becomes a relativistic matter. There is nothing wrong with this *per se*, but it presents a problem. Does David communicate to other speakers his opinion about the non-existence of color? Unless they discuss matters in a philosophical vein, we can assume not. Thus ordinary speakers will assume that when David utters the sentence "The grass is green" he is expressing a belief. Of course, David could avoid this by employing some of the standard devices for indicating the withdrawal of assertoric force. He could precede his color claims by something equivalent to "Once upon a time..."; he could utter them in a sarcastic tone of voice, or in the subjunctive mood; at a pinch, he could wear a T-shirt that declares "I withhold assertoric force from color claims!" But if he does none of these things we can assume that his interlocutors will reasonably take his color claims to be color assertions. And the possibility arises that if all listeners take an utterance to be an assertion, then, regardless of the speaker's true attitude, it *is* an assertion—in which case maybe we ought not be noncognitivists about David's color discourse after all.

If to assert *p* is to express one's belief that *p*, then it may seem impossible that David could assert "The grass is green," given our assumption that he does not believe this. But this would reveal a misunderstanding of how "express" is intended here: it indicates not a causal relation, but one established by linguistic convention. When one *lies*, for example, one expresses a belief that one does not have. That is to say, one exploits the linguistic conventions that decree that when "Such-and-such" is uttered in certain circumstances (e.g., in a serious tone of voice, not as part of a play, not preceded by "Once upon a time...," etc.) then the speaker is to be taken to believe that such-and-such. Since, we are assuming, David is not employing any of the well-entrenched devices to indicate withdrawal of assertoric force, then it might be argued that his utterance satisfies the criteria for being an assertion. And since David doesn't believe the proposition in question, then, according to this line of thinking, his alleged assertion that the grass is green looks suspiciously like a *lie*.

It would be nice to avoid the conclusion that fictionalists are liars. Let me offer two responses. First, the term "lie" is a bit steep for the situation described. David, after all, doesn't intend to deceive anyone when he utters "The grass is green." He has no malevolent agenda. He remains disposed to admit his non-belief in colors if anyone

wishes to pursue the philosophical point—it is just that such a cerebral turn is inappropriate for 99% of conversations. Though David and his interlocutors may not be on quite the same wavelength when they discuss the color of things, no harm comes of it. If "the truth about David" were to become widely known, then ordinary people may be puzzled or amused at so esoteric an idea as that the world is not colored, but it seems unlikely that they would feel annoyed at having been duped. These comments can be interpreted in either of two ways—I don't mind which: (A) expressing the belief that *p* while not believing that *p* is a necessary but not sufficient condition for lying; or (B) expressing the belief that *p* while not believing that *p* may be a sufficient condition for lying, but lying need not warrant criticism.

The second response is to move attention away from the "lone fictionalist," and remind ourselves that fictionalism is a proposed response to the question of what *we* could do if faced with an error theory concerning a hitherto fully-endorsed discourse. Fictionalism may be a stable and viable strategy for a group, even if there are some unsettling aspects of it as an individual stance. A group may have a convention in place that when a certain subject matter is entered into, there is a withdrawal of ordinary conversational force. The question of how such conventions get established and passed on is an intriguing one. Consider the murky origins of the convention of sarcasm, for example. Who decided that a certain tone of voice would act as a kind of derogatory negation of manifest content? We employ the convention without even thinking of it as "a convention"; we do not need to be explicitly taught sarcasm as children, we would have trouble articulating exactly how it works if asked to explain. The convention can also withstand the existence of a sizable number of people in the population who seem oblivious of its existence.

When fictionalism is presented in this light—as a proposal for how a *group* might respond to an error theory—we see just how "revolutionary" are the theory's aspirations. Whether such a radically prescriptive spirit is seen as simply preposterous depends on how we conceive of our philosophical objectives. Do I really expect that ordinary speakers will adjust their attitude towards a problematic discourse? Of course not. Ordinary speakers will carry on doing whatever they please. Most of them believe in ghosts, miracles, astrology and alien abductions. As philosophers writing against such silly beliefs we conceive of ourselves as correcting erroneous thought—of encouraging people to drop their false beliefs and adopt true ones—but we should not seriously expect to succeed! Revolutionary fictionalism is hardly more ambitious in its prescriptive spirit than this.

5. THE VALUE OF MORALITY

With a basic theory of fictionalism now on the table, we can turn, finally, to *moral* fictionalism. Suppose that a moral error theory is the case—or at least suppose that a group of people has become convinced of this—what should they do with their faulty moral talk? The conclusion that they should just abolish it, that it should go the way of

witch discourse and phlogiston discourse, is certainly a tempting possibility, and may, for all I say here, turn out to be the correct response. But fictionalism shows us that it is not the *only* response; it is at least possible that they may reasonably elect to maintain moral discourse as a fiction. What they need to perform is a cost-benefit analysis. Let us suppose, firstly, that the option of carrying on *believing* in morality is closed to them. They have seen the cat out of the bag and they cannot believe otherwise. Even if they *could* somehow bring themselves to sincerely "forget" that they ever read Mackie's book (for example), surely to embark on such a course is likely to bring negative consequences. I will assume without presenting any arguments that these consequences are sufficiently detrimental as to place this option beyond contention.

Similarly, I will not give serious consideration to the proposal we might call "propagandism": that *some* people may be "in the know" about the moral error theory while, for the greater good, keeping it quiet and encouraging the *hoi polloi* to continue with their sincere (false) moral beliefs. Such a situation really would amount to the promulgation of manipulative lies, which, I will assume, leads ultimately to no good. Here I agree with Richard Garner, commenting on Plato's state policy of deception in the *Republic*: "If the members of any society should come to believe Socrates' fable [the "myth of the metals"], *or any similarly fabricated radical fiction*, the result would be a very confused group of people, unsure of what to believe, and unable to trust their normal belief-producing mechanisms. It is not wise to risk having a society of epistemological wrecks in order to achieve some projected good through massive deception" (1993: 96).

Two options remain as contenders in the cost-benefit analysis: abolitionism (or we may call it "eliminativism") and fictionalism. For moral fictionalism to be viable it must win this pragmatic comparison. It is not required that taking a fictional stance towards moral discourse will supply *all* the benefits that came with sincere moral belief. It can be conceded up front that the pragmatically optimal situation for a group of people to be in is to have the attitude of sincere belief towards moral matters. But it must also be grasped that having a doxastic policy concordant with *critical inquiry* is almost guaranteed to serve better in practical terms for a group than any other policy. We are imagining a group of people whose careful pursuit of truth has overthrown their moral beliefs. Perhaps such people correctly recognize that they were happier and better off before the pursuit brought them so far, but there is now no going back, and to sacrifice the value of critical inquiry would be disastrous.

In order to assess who might win this two horse race, we must ask the question "What is the value of morality?" Unless we roughly know the answer we can have no idea of what costs its abolition may incur. Let us at first put fictionalism aside, and address the question of the value of morality *when it is believed*. We may then assume that this is a benefit that, *ceteris paribus*, will be lost if a group were to abolish morality, which puts us in a position to ask (in the next section of this paper) whether their adopting a fictionalist stance would allow them to avoid some of those losses.

The popular thought that without morality all hell would break loose in human society is a naive one. Across a vast range of situations we all have perfectly good *prudential* reasons for continuing to act in cooperative ways with our fellow humans. In many

situations reciprocal and cooperative relationships bring ongoing rewards to all parties, and do so *a fortiori* when defective behaviors are punished. When, in addition, we factor in the benefits of having a good reputation—a reputation that is based on past performance—then cooperative dispositions can easily out-compete hurtful dispositions on purely egoistic grounds.

To an individual who asks why she should not cheat her fellows if she thinks that she can get away with it, Hobbes long ago provided one kind of answer: because the punishment-enforcing power is very powerful indeed.[6] This answer is developed and supplemented by Hume, who speaks of knaves "betrayed by their own maxims; and while they purpose to cheat with moderation and secrecy, a tempting incident occurs, nature is frail, and they give into the snare; whence they can never extricate themselves, without a total loss of reputation, and the forfeiture of all trust and confidence with mankind" (1751/1983: 82). First, the knave misses out on benefits that by their very nature cannot be gained through defection: "Inward peace of mind, consciousness of integrity, a satisfactory review of [her] own conduct" (1751/1983: 82)—advantages that are constituted by a disposition not to cheat one's fellows. Moreover, the knave will lose these benefits for comparatively trivial gains ("the feverish, empty amusements of luxury and expence"). Third, knaves will be epistemically fallible, and might think that they can get away with something when in fact they will be caught and punished. Fourth, since knaves have on their minds the possibility of cheating whenever they are confident of evading detection, they are likely to be tempted to cheat in situations where the chances of evading detection are less than certain, thus, again, risking severe punishment.

One result we can draw from Hobbes and Hume is that a person may have many reasons for acting in accordance with a moral requirement: the fear of punishment, the desire for an ongoing beneficial relationship, the motivation to maintain a good reputation, the simple fact that one on the whole *likes* one's fellows, that one has been brought up such that acting otherwise makes one feel rotten—all these being solid prudential reasons—plus the moral requirement to act. To subtract the last one leaves the others still very much in play. But if this is so, then what useful role does the last kind of consideration play at all? To answer this it is worth underlining the reference to *temptation* in Hume's answer to the sensible knave. Merely to believe of some action "This is the one that is in my long-term best interests" simply doesn't do the job. Most of us know this from personal experience, but there is abundant empirical evidence available

---

[6] Given that it is in an individual's interests to engage in mutually beneficial contracts, it will be in her interests to support a social system wherein contractual compliance is enforced. Of course, for any individual the *optimal* scheme is if her neighbors are forced to comply and she alone is able to break contracts and evade punishment—but such an arrangement, we may assume, is not an available option. When the only options concern a *non-discriminating* police force, it will be to each individual's interests to choose the maximally vigilant sovereign power. That way a given individual will have to forego the benefits of cheating others, but stands the best chance of avoiding the proportionally greater costs of being cheated (bearing in mind that the disadvantages of having one's throat cut are far greater than any advantages that may accrue from cutting another's throat).

for the dubious (see Ainslie 1975; Schelling 1980; Elster 1984, 1985). Because short-term profit is tangible and present whereas long-term profit is distant and faint, the lure of the immediate may subvert the agent's ability to deliberate properly so as to obtain a valuable delayed benefit, leading him to "rationalize" a poor choice. Hobbes lamented this "perverse desire for present profit" (1642/1983: 72)—something which Hume blamed for "all dissoluteness and disorder, repentance and misery" (1751/1983: 55), adding that a person should embrace "any expedient, by which he may impose a restraint upon himself, and guard against this weakness" (1739/1978: 536-7[7]). Let me hypothesize that an important value of moral beliefs is that they function as just such an expedient: supplementing and reinforcing the outputs of prudential reasoning. When a person believes that the valued action is *morally* required—that it *must* be performed whether he likes it or not—then the possibilities for rationalization diminish. If a person believes the action to be required by an authority from which he cannot escape, if he imbues it with a "must-be-doneness" (the categorical element of morality that Mackie found so troublesome), if he believes that in not performing he will not merely frustrate himself, but will become reprehensible and deserving of disapprobation—then he is more likely to perform the action. The distinctive value of categorical imperatives is that they silence calculation, which is a valuable thing when interfering forces can so easily hijack our prudential calculations. In this manner, moral beliefs function to bolster self-control against practical irrationality.

   I would not go so far as to claim that this is *the* value of moral belief, or even the most important benefit—but the argument requires only that we locate one general and reliable source of practical value. This suffices to show why a moral error theorist should hesitate before embracing abolitionism, for it reveals a practical cost that would be incurred on that path. (If there are other sources of practical benefit brought by moral beliefs, then the costs of abolitionism are even higher.) The crucial question, then, is whether some of the costs may be avoided by taking a fictionalist stance towards morality—whether the practical benefits of moral belief may still be gained by an attitude that falls short of belief. On the face of it, it seems unlikely. How can a fiction have the kind of practical impact—moreover, the kind of practical *authority*—that confers on moral belief its instrumental value? This is the major reason that moral fictionalism seems troubling in a way that color fictionalism does not: it seems implausible that a mere fiction could or should have such practical influence on important real-life decisions. In what remains of this paper let me try to assuage this reasonable doubt.

6. MORAL FICTIONALISM

First let me reiterate the caution already noted: that it is not incumbent on the moral fictionalist to argue that taking a fictional attitude towards morality makes *no* difference,

---

[7] I have altered Hume's text from the first person to the third person singular.

or that morality as a fiction will supply *all* the practical benefits of a believed morality. A background assumption is that the arguments for moral error theory have put the option of a *believed* morality out of the running, so the only comparison in which we are interested is between fictionalism and abolitionism. The fictionalist wins the argument if she shows that there is *some* benefit to be had from keeping moral discourse as a fiction that would be lost (with no compensating gain) by eliminating moral discourse entirely.

In the previous section I argued that an important practical benefit to the individual of having moral beliefs is that they will serve as a bulwark against weakness of will—silencing certain kinds of vulnerable calculation, and thus blocking the temporary re-evaluation of outcomes that is characteristic of short-sighted rationalization. So our task is limited to addressing the question of whether a "mere fiction" could also provide a similar benefit.

A quick argument to show that a positive answer is within reach begins by noting that engagement with fiction can affect our emotional states. This view is not without detractors: Kendall Walton, for example, has argued that fictions do not produce real emotions, but rather make-believe emotions (see his 1978, 1990[8]). But this is a terribly counter-intuitive view, which I am confident is incorrect. All the empirical evidence supports commonsense on this matter: watching movies, reading novels, or simply engaging one's imagination can produce real episodes of fear, sadness, disgust, anger, and so on. (One explanation is, in the words of two eminent psychologists, simply "that the cognitive evaluations that engender emotions are sufficiently crude that they contain no reality check" (Johnson-Laird and Oatley 2000: 465); alternatively, one may think that the human tendency to enjoy fictional engagement served some adaptive purpose in the ancestral environment.[9]) To this premise we can add the truism that emotional states can affect motivations, and thus behavior. Of course, the emotions arising from fictions do not necessarily affect behavior in the same manner as emotions arising in response to beliefs: the fear of fictional vampires is consistent with my sitting eating popcorn, whereas fear of vampires in which I *believed* would result in purchasing wooden stakes and a lot of garlic. But it does not follow that the emotions arising from engagement with fiction are "motivationally inert." Reading *Anna Karenina* may encourage a person to abandon a doomed love affair; watching *The Blair Witch Project* may lead one to cancel the planned camping trip in the woods. Needless to say, these aren't the kind of beneficial behavioral responses that the moral fictionalist is seeking, but they at least show that the causal links between involvement with a fiction and action are undeniably in place.

---

[8] Others who reject the view that we have genuine emotions in response to fiction include Kenny (1964) and Budd (1985).

[9] The latter hypothesis gains support over the former when one considers that in fictional encounters people enjoy and seek out emotions that they otherwise generally avoid (fear, sadness, etc.). The evolutionary hypothesis holds that the capacity to engage with fiction and make-believe is a kind of "safe training" for real life risks and opportunities. Natural selection makes the accompanying emotions enjoyable in order to motivate the activity (for the same reason as it makes eating and sex enjoyable). See Steen and Owen (2001).

Let us turn our sights more directly on the question of how a person combats weakness of will. Suppose I am determined to exercise regularly, after a lifetime of lethargy, but find myself succumbing to temptation. An effective strategy will be for me to lay down a strong and authoritative rule: *I must do fifty sit-ups every day, no less*. I am attempting to form a habit, and habits are formed—and, for the doggedly weak of will, maintained—by strictness and overcompensation. Perhaps in truth it doesn't much matter that I do fifty sit-ups every day, so long as I do more-or-less fifty on most days. But by allowing myself the occasional lapse, by giving myself permission *sometimes* to stray from the routine, I pave the way for akratic sabotage of my calculations—I threaten even my doing more-or-less fifty sit-ups on most days. I do better if I encourage myself to think in terms of fifty daily sit-ups as a non-negotiable value, as something I *must* do if I am ever to get fit.

However, to believe sincerely that fifty daily sit-ups are needed in order for me to achieve fitness is to have a false belief (we'll assume), the holding of which will require other compensating false beliefs. If it is true that *more-or-less* fifty sit-ups *nearly* every day is sufficient for health, then that is what I ought to believe. On the other hand, to *pay attention* to this belief exposes me to self-subversion—a slippery slope to inactivity. This is precisely a case where my best interests are served by rehearsing thoughts that are false, and that I know are false, in order to fend off my own weaknesses. But in order to get the benefit from this strategy there is no necessity that I *believe* the thoughts, or attempt to justify them as true when placed in a philosophically critical context. While doing my sit-ups I think to myself "Must...do...fifty!" but if, on some other occasion, you ask me whether I really *must* do fifty, then I will say "No, sometimes forty would suffice."

Human motivation is often aroused more effectively by mental images than by careful calculation. Hume uses the example of a drunkard "who has seen his companion die of a debauch, and dreads a like accident for himself: But as the memory of it decays away by degrees, his former security returns, and the danger seems less certain and real" (1739/1978: 144). Hume's point is that humans put weight on near, recent and concrete evidence, though there is no rational justification for our doing so. We can imagine the drunkard being presented with impressive statistics on the probabilities of alcoholics suffering an unpleasant end, but remaining quite unmoved; yet one friend dies and he becomes a teetotaler (at least for a while). It's not that he disbelieved the statistics, and the death of the friend need not alter his beliefs about how likely he is to suffer a similar fate, but the "tangibility" of the one death has, in Hume's words, "a superior influence on the judgment, as well as on the passions" (1739/1978: 143-4).

If the drunkard has decided that his long-term interests are best served by abstinence, what strategy should he pursue to that end? He should read the statistics, yes, but—perhaps even more importantly—he should attempt to keep the image of his dying friend vivid. He does still better if he can relate that image to his own plight, if he thinks: "If I drink, that's what will happen to me." Now that proposition is false. What is true is something like "If I drink, there's a 10% chance [say] of that happening to me." But *that* thought looks dangerous. He does better with the stronger: "If I drink, that's what *will* happen to me." Yet does he, need he, *believe* this? No: he need not believe it in order for

it to affect his actions in the desirable way, and, moreover, he *ought not* believe it because it is false.

Hume's view that decisions are influenced by the "tangibility" of how information is presented receives ample empirical support. In a large scale survey conducted on doctors' attitudes towards smoking in the 1970s, it was noted that smoking had dropped most dramatically in chest physicians and radiologists—those who had been exposed to the effects of the activity—while other types of doctor, though no doubt aware of the statistics, were much less moved (Borgida and Nisbett 1977). "Tangibility" also affects the willingness of a person to enter into a mutually beneficial cooperative relationship. It has been shown that pairs of people playing iterated Prisoner Dilemma games will be much more likely to develop a cooperative strategy if the information concerning how the other player acted in the previous round is conveyed by a written note passed through a slot, as opposed to one of two small lights being activated (Enzle et al. 1975). The same information is disclosed by either means, but one form is (in a way that's difficult to articulate) more "concrete," more "palpable," than the other, according it a greater influence in deliberations.

In another study of how people play Prisoner's Dilemma games it was shown that if, while sitting in the waiting room prior to playing the game, a person overhears a (fake) radio news item about an act of sacrifice (such as the donation of a kidney) then the person will be much more likely to adopt a cooperative strategy in the subsequent game (Hornstein et al. 1975). By comparison, a radio story presenting violence and nastiness will encourage listeners subsequently to adopt a non-cooperative strategy. It is possible that a "nice" news story affects the person's mood in a way conducive to cooperation, or perhaps it places in his short-term memory a kind of role model, or temporarily makes certain features of the real world appear more salient in deliberations. However it works, it is pretty clear that an engagement with a fictional story (as opposed to an apparent news item) may have a similar affect (though, to my knowledge, the obvious experiment has not been done).

Though these studies may be unfamiliar, what they reveal should hardly come as a surprise. The whole advertising industry (with which we are all far more familiar than we would wish) operates on the assumption that heavily exaggerated, idealized, and fictional images and narratives can influence real choice. We are shown an image of an absurdly happy family living in an eternally sunny world, and the basis of their rapture, we are encouraged to think, is the cereal that sits in the center of the breakfast table. Do we believe such garbage? Not for a second.[10] Do we, nevertheless, go out and spend our hard-earned money on that cereal? Much as we would like to deny it, masses of empirical research shows that we do.

One may object that choosing breakfast cereals hardly compares to moral decision-making, but it would be naive to deny that the same advertising strategies can encourage

---

[10] In a study conducted in 1971, it was shown that only 12% of sixth graders believed that television commercials told the truth all or most of the time (Lyle et al. 1971).

us to give to charity, vote for a president, support a bombing campaign, or sign up to join the armed forces. That engagement with fiction can influence our deliberations over the most weighty decisions is beyond question. What is perhaps unusual about the situation of the fictionalist, and which requires more discussion, is the proposal that the action-guiding fiction be in some manner self-generated.

7. MORAL FICTIONALISM AS A PRECOMMITMENT

Sometimes, when on a long airplane flight, I succumb to weakness of will and eat all the awful inflight food that I had promised myself I wouldn't eat. It happens because I am trapped and bored with the food right in front of me for a long time. In order to avoid this I have developed a strategy for resisting my own imprudence. If I have decided that I really don't want to eat that slice of cheesecake, but suspect that I won't be able to resist picking at it until it's all gone (despite its tasting of plastic), I smear some gravy on top of it. (It raises the eyebrows of the person sitting next to me, but certainly ensures that I won't eat the cheesecake.) In doing this I am, in a very unglamorous way, following the example of Odysseus when he had himself bound to the mast of his ship so as not to give in to the song of the sirens. The circumstance in which he made that decision was one in which he was free of temptation, but he was shrewd enough to anticipate the overthrow of control. Such strategies for combating weakness of will John Elster calls "precommitments" (1984: 37ff).

   The decision to adopt morality as a fiction is best thought of as a kind of precommitment. It is not being suggested that someone enters a shop, is tempted to steal, decides to adopt morality as a fiction, and thus sustains her prudent though faltering decision not to steal. Rather, the resolution to accept the moral point of view is something that occurred in the person's past, and is now an accustomed way of thinking. Its role is that when entering a shop the possibility of stealing doesn't even enter her mind. If a knave were to say to her "Why not steal?" she would answer without hesitation "No!—Stealing is wrong." What goes through her mind may be exactly the same as what goes through the mind of the sincere moral believer—it need not "feel" like make-believe at all (and thus it may have the same influence on behavior as a belief). The difference between the two need only be a *disposition* that the fictionalist has (though is not paying attention to): the disposition to deny that anything is really morally wrong, when placed in her most critical context.[11]

---

[11] It is worth reminding ourselves that "critical context" is a term of art, and in other vernacular senses of the phrase it is those times when the person is immersed in the fiction that involve more critical thinking. Working out the plot of a complex novel, for example, may involve a great deal of careful thinking, whereas the thought "It's all just a fiction" is a simple matter. Nevertheless, in the sense defined, the latter is the more "critical context" since it questions and challenges the world of the novel. In the same way, though a moral fictionalist will reject moral claims when doing metaethics, this is perfectly consistent with her employment of the moral fiction at other times involving an enormous amount of critical deliberation

But what if the knave carries on: "But in all seriousness, taking into account philosophical issues, bearing in mind John Mackie's arguments—*why not steal*?" Then, *ex hypothesi*, our fictionalist will "step out" and admit that there is nothing morally wrong with stealing. So does she then stuff her pockets? No! For she still has all those Hobbesian and Humean reasons to refrain from stealing. It is no part of the argument of this paper that moral thinking should be followed if it prescribes actions that we do not have good reasons for performing independently of moral considerations. One would deny this at the price of allowing that morality may serve no purpose to the individual at all. If we embrace the view that a believed morality is useful to the individual, then we must be employing some non-moral standard by which to make this assessment. If (as seems correct) an individual's believing that some available action is morally required increases the probability of his performing that action, then it seems plausible to assume that the usefulness to an individual of moral belief lies at least in part in its increasing the probability of his performing those actions that he judges he morally ought. From these assumptions it follows that such actions were useful to him anyway—i.e., that he had a non-moral reason for performing them.

The idea of the precommitment to the moral fiction being a conscious choice that someone makes is an artificial idealization. (In this it differs from pouring gravy on cheesecake.) It is more likely that a person is simply brought up to think in moral terms; the precommitment is put in place by parents. In childhood such prescriptions may be presented and accepted as items of belief (it is not implausible to hold that the best way to encourage prudent habits is to tell children a few white lies); thus thinking of certain types of action as "morally right" and others as "morally wrong" becomes natural and ingrained. Later, when a broader and more sophisticated understanding is possible, the person may come to see how philosophically troubling is the idea that there really are actions that people *must* perform, irrespective of whether they wish to, regardless of whether it suits their ends—and if convinced by such arguments she becomes a moral error theorist. But these patterns of thought might be now so deeply embedded that in everyday life she carries on employing them—she finds it convenient and effective to do so, and finds that dropping them leaves her feeling vulnerable to temptations which, if pursued, she judges likely to lead to regret. There is, besides, a practical value to be gained simply from the convenience of carrying on in the manner to which she has grown accustomed. She doesn't cease to be a moral error theorist, but she becomes, in addition, a moral fictionalist.

There are no doubt other ways of combating weakness of will. Perhaps some strategies are, taken alone, more effective than adopting a fictive attitude towards the "must-be-doneness" of the optimal option. All that the present argument requires is that adopting a fictionalist stance would provide *some* help in strengthening resolve in addition to any other effective strategies. (Bear in mind also that I am not arguing that acting as a bulwark against temptation is the *only* value of morality, so even if my arguments

---

and careful calculation.

concerning the contribution that a moral fiction may make in this respect fail to convince, moral fictionalism does not thereby fall flat.) In fact, the preceding argument entails that there is at least one other effective way of combating weakness of will. Why, one might start out wondering, isn't the decision to adopt morality as a fiction subject to weakness of will? If the presence of the shiny money within reach is likely to tempt one to grab it, ignoring the voice of prudence that is warning that this will lead to no good end, then why won't the same lure of short-term profit also incite the immediate abandonment of the moral fiction? The answer I gave is that the moral fiction is a precommitment that can exclude from practical deliberation the entertainment of certain options: all going well, the fictional attitude blocks the temptation to steal from even arising (just as does, all going well, sincere moral belief). But if this answer is reasonable here, then isn't the same kind of answer, the same kind of prudence-reinforcing strategy, available without any fictionalizing entering the picture at all? Why can't a person simply have the precommitment not to steal (plus a precommitment to keep promises, to refrain from initiating violence, etc.)?

It is not clear what it means simply to have "a precommitment not to steal (etc.)." Perhaps it means a *habit* of not stealing, such that a person is brought up so the thought of stealing simply doesn't enter his mind. Or perhaps it means a habit of feeling sympathy for fellow humans, such that the prospect of harming them by stealing from them motivates one to refrain from doing so. But though encouraging such habits may be a very good way of fortifying clear-headed instrumental reasoning (which, for Hobbesian and Humean reasons, generally comes down against stealing), my contention is that they would work even more effectively if supplemented with *moralized* thought.

Suppose that a person with no moralized thinking (neither as belief nor fiction) were, despite his voice of prudence properly counseling otherwise, for some reason to steal. Let's assume that he has in place a habit of not stealing, and a habit of feeling sympathy for others' suffering, but nevertheless these habits were not on this occasion strong enough to withstand the temptation of short-term profit. How does he now feel? The fact that he has broken a habit may surprise him. The fact that he has hurt someone that he didn't want to hurt may cause him disappointment and distress. But the important thing is that he can feel no *guilt*, for guilt requires the thought that one has done something *wrong*. With no moral concepts in play, this person does not have access to the thought that he *deserves* to be punished for his action; he regrets, but he cannot repent. His active sympathy may prompt in him a desire to alleviate the victim's suffering (he may even feel a desire to return the stolen goods), but since he has no thought that he *must* do something to make amends, were he to become distracted by other matters, such that his sympathy for the victim fades, then there is nothing to propel his deliberations back to the resolution that "something must be done." In the end, he has just done something out of character that he wishes he hadn't done. "Sympathy," J.Q. Wilson once wrote, "is a fragile and evanescent emotion. It is easily aroused but quickly forgotten; when remembered but not acted upon, its failure to produce action is easily rationalized. The sight of a lost dog or a wounded fledgling can upset us greatly even though we know that the woods are filled with lost and injured animals" (1993: 50).

By comparison, the person who can "moralize" her thoughts (either as belief or fiction) will feel differently if on occasion she succumbs to temptation. She can tell herself that she has done something *wrong*, that her action was *unfair*, that she must make amends, that she not only has risked punishment, but also *deserves* it. (In addition, she can judge that other felons deserve punishment too—a thought that was unavailable to our previous non-moral agent.) The fact that these more robust forms of self-recrimination are available to the moral thinker when she does steal strongly suggests that when she is behaving herself her motivation not to steal is more reliable and steadfast than that of her non-moral counterpart. Her deliberations and justifications do not end in the thought "Well, I just don't want to do that," but rather the more vivid and non-negotiable "That would be wrong."

Of course, what ultimately determines whether a person will refrain from stealing is the strength of the desire not to steal compared with the desire to do so. The claim is that the thought "That would be wrong" plays a role in desire-formation and is likely to *strengthen* any desire against stealing that one has as the result of any "non-moralized" habit. It is true that this thought as a fictive judgment may not play as robust a role in an agent's desiderative life as the thought as a belief, but so long as it reliably pulls *some* weight—so long, that is, as the fictionalist reliably has a pragmatic advantage over the moral eliminativist—then the error theorist is justified in keeping moral discourse as a "useful fiction."

8. CONCLUSION

The advice "Maintain moral discourse as a fiction" is not intended to apply necessarily to any agent in any circumstances. It would be unreasonable to expect that it should, especially since the legitimacy of any more authoritative kind of prescription—e.g., to the effect that one *must* adopt the moral fiction, irrespective of one's ends or interests—is likely to have been rejected in the prior argument for a moral error theory (the details of which argument this paper has, for obvious reasons, skirted). It is enough if it turns out to be good advice for *us now*: people who are prone to temptation, epistemically fallible, and familiar with moral thinking. I have offered an argument in support of its being good advice, but of course ultimately it is an empirical matter which depends on the ability to assess far-fetched counterfactuals, and I am the first to admit that it may all turn out to be mistaken. It is possible that moral fictionalism deserves a place on the menu of metaethical options while the prescription urged by those of us on the "revolutionary wing" of the theory remains poor advice.

Since this paper has presented no arguments in favor of a moral error theory, discussing the prospects of moral fictionalism may seem premature. I agree that the preferred strategy must always be to do our utmost to show that moral discourse is not really flawed at all—and I dare say that nearly all readers believe this battle still to be worth fighting. But the viability of moral fictionalism should be of more than academic interest even to those who are not error theorists, for I suspect that those eager to

repudiate the error theoretic position often derive their concern in part from worries about what might *happen* if the theory were to become widely accepted as true. It is viewed not merely as counter-intuitive, but as a genuinely threatening and pernicious doctrine. David Brink, for example, once suggested that we should learn to live with whatever "metaphysical queerness" is entailed by moral realism if the only alternative "would undermine the nature of existing normative practices" (1989: 173). But if this kind of concern is unjustified—as the possibility of moral fictionalism suggests it may be—then the motivation for resisting a moral error theory is in need of re-examination.[12]

---

[12] Thanks to Stuart Brock, Fred Kroon, and Jerry Vision for useful feedback.

References:

AINSLIE, G., (1975). Impulsiveness and impulse control. *Psychological Bulletin* 82: 463-96.

BARON-COHEN, S., (1987). Autism and symbolic play. *British Journal of Developmental Psycho*logy 5: 139-148.

BORGIDA, E. AND NISBETT, R.E., (1977). The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology* 7: 258-71.

BRINK, D., (1989). *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

BUDD, M., (1985). *Music and the Emotions*. London: Routledge and Kegan Paul.

BURGESS, J., (1983). Why I am not a nominalist. *Notre Dame Journal of Formal Logic* 24: 93-105.

BURGESS, J. AND ROSEN, G., (1997). *A Subject with No Object*. Oxford: Clarendon Press.

ELSTER, J., (1984). *Ulysses and the Sirens*. Cambridge: Cambridge University Press.

ELSTER, J., (1985). Weakness of will and the free-rider problem. *Economics and Philosophy* 1: 231-65.

ENZLE, M.E., HANSEN, R.D., AND LOWE, C.A., (1975). Humanizing the mixed-motive paradigm: methodological implications from attribution theory. *Simulation and Games* 6: 151-65.

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

HOBBES, T., [1642] (1983). *De Cive*. Oxford: Clarendon Press.

HORNSTEIN, H.A., LAKIND, E., FRANKEL, E., AND MANNE, S., (1975). Effects of knowledge about remote social events on prosocial behaviour, social conception, and mood. *Journal of Personality and Social Psychology* 32: 1038-46.

HUME, D., [1739] (1978). *A Treatise of Human Nature*. Oxford: Clarendon Press.

HUME, D., [1751] (1983). *Enquiry Concerning the Principles of Morals*. Cambridge,

MA.: Hackett Publishing Company.

JARROLD, C., BOUCHER, J., AND SMITH, P.K., (1996). Generativity deficits in pretend play in autism. *British Journal of Developmental Psychology* 14: 275-300.

JOHNSON-LAIRD, P. AND OATLEY, K., (2000). Cognitive and social construction in emotions. *In*: M. Lewis and J. Haviland-Jones, eds. *Handbook of the Emotions*. 2nd ed. NY: Guilford Press: 458-75.

KENNY, A., (1964). *Action, Emotion and the Will*. London: Routledge and Kegan Paul.

LEWIS, D., (1978). Truth in fiction. *American Philosophical Quarterly* 15: 37-46.

LYLE, J., AND HOFFMAN, H., (1971). Children's use of television and other media. *In*: J.P. Murray, E.A. Robinson, and G.A. Comstock, eds. *Television and Social Behavior*. Volume 4, Rockville, MD: National Institutes of Health: 129-256.

MACKIE, J., (1977). *Ethics: Inventing Right and Wrong*. NY: Penguin Books.

QUINE, W.V., (1960). *Word and Object*. Cambridge, MA.: MIT Press.

SCHELLING, T., (1980). The intimate contest for self-command. *The Public Interest* 60: 94-118.

STANLEY, J., (2001). Hermeneutic fictionalism. *Midwest Studies in Philosophy 25: Figurative Language*: 36-71.

STEEN, F.F. AND OWEN, S.A., (2001). Evolution's pedagogy: an adaptationist model of pretense and entertainment. *Journal of Cognition and Culture* 1: 289-321.

VISION, G., (1994). Fiction and fictionalist reductions. *Pacific Philosophical Quarterly* 74: 150-74.

WALTON, K., (1978). Fearing fictions. *Journal of Philosophy* 75: 5-27.

WALTON, K., (1990). *Mimesis and Make-Believe*. Cambridge: Harvard University Press.

WILSON, J.Q., (1993). *The Moral Sense*. NY: Free Press.