Psychological Fictionalism, and the Threat of Fictionalist Suicide

1. Introduction

"Eliminativism" is an ambiguous term. When applied to psychological entities, like beliefs and desires, "eliminativism" (as in "eliminative materialism") usually denotes the view that the entities widely referred to do not exist. This position—championed most famously by Paul Churchland (1979, 1981) and Patricia Churchland (1986)—is an error- theory regarding psychological entities. In this context, the verb "eliminate" denotes the act of banishing something from our *ontology*; we realize that there is no place for beliefs and desires in our mature world view. It is not the intention of this paper to assess the truth of psychological eliminativism; despite its being a radical and fairly unpopular view, I will simply assume that there are respectable arguments in its favor. (For defense of the view, beyond the Churchlands, see Rorty [1970]; Stich [1983]; Ramsey et al. [1990]; Taylor [1994].)[1]

In other contexts, by contrast, eliminativism is a theoretical option one may or may not choose to adopt after one has embraced the error theory. Here, the verb "eliminate" denotes the act of banishing something from our *language*; we decide that most uses of the terms "belief," "desire," etc., should be dropped.

Thus one might be a psychological eliminativist twice over: one might decide that there are no such things as beliefs and desires, and then one might decide that most uses of these terms should be jettisoned. Let us call these views "ontological eliminativism" and "linguistic eliminativism," respectively. It is so natural to assume that the latter form of eliminativism should accompany the former that many of the classic statements of eliminative materialism fudge the matter.[2] Yet the two can come apart. One can be an eliminativist in the first sense—endorsing an error theory for certain classes of psychological entity—yet resist eliminativism

in the second sense—allowing that talk of these entities should be maintained. This combination of views, which may be called "psychological fictionalism," will be explored in this paper. I will delineate a number of different kinds of psychological fictionalism. My goal is not to advocate any of these theories, though the final section of this paper is devoted to rebutting a charge that may be leveled at psychological fictionalism: the threat of fictionalist suicide.

As a preliminary, let me explain what I mean by saying that according to linguistic eliminativism *most* uses of the offending term should be dropped. The Churchlands liken folk psychology to talk of vitalism (Paul Churchland 1981, 71, 89; Patricia Churchland 1981, 100–101)—something about which one should be an error theorist. Yet nobody claims that *all* appearances of the phrase "vitalistic life force" must be dropped from our discourse—even true sentences like "There is no such thing as vitalistic life force" and "Pasteur designed experiments to test whether there is a vitalistic life force." Rather, the linguistic eliminativist about vitalism argues for the abolition of all utterances that commit the speaker to the existence of vitalistic force. Assertion of these sentences do not carry this commitment. Nor does the mere *utterance* of the sentence "Vitalistic life force exists" commit one to the existence of vitalistic force. One might, for example, utter this sentence without assertoric force, if asked to provide an example of a false four-word sentence; or one might say it as a joke, or as a line in a play. As W.V. Quine once put it: "The parent who tells the Cinderella story is no more committed to admitting a fairy godmother and a pumpkin coach into his own ontology than to admitting the story as true" (Quine 1961, 103). The linguistic eliminativist about psychological entities argues for the abolition of all utterances that commit the speaker to the existence of certain psychological entities.

Characterizing linguistic eliminativism in this way does not succeed in distinguishing it from the fictionalist alternative, for the fictionalist about psychological entities also shuns utterances that commit the speaker to the existence of these entities. Where the theories differ is that the fictionalist holds that all those utterances that one would ordinarily think of as committing the speaker to psychological entities in fact do not (or need not) do so, and thus there is no pressure for their abolition. So, for example, both the eliminativist and fictionalist agree that there is nothing amiss with uttering the sentence "Vitalistic life force exists" in the course

of reciting a line of a play, for such a context makes no ontological commitment. Where the difference arises is regarding all those conversational contexts where one would usually assume that the utterance of this sentence does evince a commitment to the existence of vitalistic force: that is, seemingly assertoric talk where there is no acting, no joking, no sarcasm, no quoting, etc. The eliminativist takes such talk at face value and argues for the abolition of any utterance of "Vitalistic life force exists" from such contexts. Such an attitude may be thought of as the orthodox or commonsensical response. (It is, after all, the attitude we do take toward discourse about vitalism.) But the fictionalist takes an unorthodox route, arguing that the commitment here is only apparent. Those contexts that we would ordinarily think of as assertoric conversation are in fact, the fictionalist thinks, not dissimilar from reciting the lines of a play: they do not ontologically commit the speaker.

2. Hermeneutic Versus Revolutionary Fictionalism

One kind of fictionalist—the hermeneutic fictionalist—presents the theory as a descriptive analysis of actual linguistic practice. Regarding psychological entities like beliefs and desires, such a fictionalist will argue that though an ordinary utterance of, say, "Mary believes that *p*" *appears* to commit the speaker to the existence of beliefs, in fact it doesn't. There is something about the domain of psychological entities, it is argued, that means that utterances involving their reference are best interpreted as governed by commitment-nullifying conventions. Just as we can nullify commitment by adopting an overtly sarcastic tone of voice, so too (it is argued) can we do so by employing certain terms (like "belief" and "desire"). Such a view counts as ontological eliminativism to the extent that it rejects the existence of beliefs and desires,[3] yet it doesn't comfortably count as an error theory. An error theory, we can assume, accuses an epistemic community of widespread *error*, but if speakers were never in the business of ontologically committing themselves to beliefs and desires when they spoke of them, then they could hardly be accused of an error. Where is the mistake in uttering the falsehood "1 + 2 = 5" if one is just kidding?

A view that allows the illustration of some of these distinctions is Gilbert Ryle's. Ryle (1949) argues that utterances like "Mary believes that *p*" function as *inference tickets*—providing license for the speaker to

move from one factual statement to another, while not themselves being factual statements. This may be construed as a form of ontological eliminativism, since it denies the existence of beliefs and desires, and it is, moreover, a hermeneutic view, since it purports to describe how psychological language actually functions. Yet it is not an error theory, inasmuch as it rejects that psychological language was ever really in the business of describing the mind in the first place, and thus could hardly be erroneously misdescribing it. Nor should Ryle's view count as a form of fictionalism, since it in no way interprets the attitude that people adopt in employing terms like "belief" and "desire" to be anything like embracing or discussing a fiction. Fictionalism is not just the negative thesis that utterances that appear to make ontological commitments do not do so, but also the positive (though vague) thesis that the role that these utterances do play is substantively similar to the role of familiar fictional discourse.[4]

Another kind of fictionalist—the revolutionary fictionalist—presents the theory as a recommendation of how we should change our linguistic practices. Such a fictionalist is an ontological eliminativist and an error theorist: accusing ordinary speakers of making false assertions when they say things like "Mary believes that p" and "John desires that q." This fictionalist argues that, in actual discourse, when people assert these sentences they really are committing themselves to the existence of (nonexistent) beliefs and desires. But whereas the linguistic eliminativist responds that these utterances must therefore be banished, the revolutionary fictionalist responds that we can alter our practices such that the utterances no longer carry the problematic commitment. In fact, the revolutionary fictionalist argues that we *should* alter our linguistic practices in this manner. We should carry on saying false things, but we should stop *asserting* them.

At this point, many questions jostle for attention. How could we take psychological discourse seriously if it lacked commitment to psychological entities? What would be the point of having such a noncommitted discourse? Why should altering the ontological commitments of a linguistic practice be superior to dropping the practice? I shan't have space to address all these natural questions here, though I shall have something to say about some of them. Rather, what I want to describe is a further distinction in how the fictionalist might describe the noncommitted practice: between cognitivist and noncognitivist versions of fictionalism. This distinction runs obliquely to the aforementioned one, allowing for cognitivist

hermeneutic fictionalism, noncognitivist hermeneutic fictionalism, cognitivist revolutionary fictionalism, and noncognitivist revolutionary fictionalism.

3. Cognitivist Versus Noncognitivist Fictionalism

The cognitivist fictionalist holds that sentences like "Mary believes that p" should be construed as containing a tacit story operator: "According to fiction F, Mary believes that p." This view counts as cognitivist in virtue of the fact that it allows that "Mary believes that p" may be asserted and may be true; in such circumstances it will be a true assertion *about* a fiction. That ontological commitment to beliefs is nevertheless removed becomes clear when we compare "According to Perrault, a fairy godmother changed a pumpkin into a carriage," the assertion of which carries commitment to the existence of neither godmothers nor transforming pumpkins.

The cognitivist fictionalist owes us an account of what the relevant "fiction F" is, for in order to assess the truth of a statement like "Mary believes that p," we will need to know what follows and what does not follow "according to fiction F."

In the case of a story by Perrault, one might think that we have a pretty good idea as to what the content of the relevant fiction is: We need merely to read the story. Yet it is not quite so straightforward, for not all things that are true according to the story are stated explicitly therein. It is not explicitly stated that Cinderella has five toes on each foot, yet it is reasonable to assume that it is true in the Perrault story that she does have five toes on each foot. (See Lewis 1978.) However, it is very difficult to assess the extent of the domain of truths that may reasonably be considered true-but-not-explicitly-stated in a fictional story. It is not explicitly stated that Cinderella has forty-six chromosomes, yet is it true in the Perrault story that she has forty-six chromosomes? The fairy tale is set in a world that seems reminiscent of 18th-century Europe, yet is it true in the Perrault story that the Roman Empire fell centuries earlier? I, for one, do not have much of an intuition on these questions, and I suspect it is because the matter is indeterminate. Still, generally speaking this indeterminacy in fictions doesn't get us into any trouble, since the indeterminacy resides in exactly the places where it doesn't matter one way or the other to the consumer of the fiction.

In the case of psychological fictionalism, the fiction in question might be called "folk psychology." This is the theory that eliminative

materialists think is false. But even if false, the theory presumably has enough content to ground "According to folk psychology. . ." claims. Some are straightforward: "According to folk psychology, beliefs exist"; "According to folk psychology, some desires are strong and some are weak." Some will be more complex: "According to folk psychology, if S desires X, and S believes that fing is the optimal means of attaining X, then this may lead to S's fing." (See Churchland 1981, 71.) But what about a claim that concerns a particular, such as "According to folk psychology, Mary believes that p"? The theory of folk psychology presumably makes no reference to the individual Mary. The sentence is, nevertheless, a permissible one. Compare the claim "According to Einsteinian physics, the light from star 66 Tauri is deflected by the Sun's gravitational field." Einstein's theory is general; its complete articulation makes no reference to particulars like "66 Tauri" or "the Sun." The claim is nevertheless reasonable in virtue of Einstein's theory *in conjunction with data concerning particulars* predicting or describing the light from a particular star being deflected by the Sun. In a similar way, the theory of folk psychology in conjunction with data concerning particulars will predict or describe certain phenomena, such as Mary's believing that p. These data will be those things that we would ordinarily take to be evidence of Mary's believing that p: her behavior, her saying so, the fact that she was just told this and didn't object, and so on. Indeed, it might be a tenet of folk psychology that certain types of things count as evidence of a person's beliefs. When we couple this tenet with the observation that Mary manifests this kind of evidence, it becomes perfectly reasonable to say "According to folk psychology, Mary believes that p."

The cognitivist fictionalist, as we have seen, translates the sentence "Mary believes that p" into something that is assertible and has truth value. One might instead translate it into something that cannot be asserted and lacks truth value—perhaps something in the optative mood, like "Let's pretend that Mary believes that p." This would be a kind of *noncognitivist* fictionalism. Another kind of noncognitivist fictionalist offers no translation at all, but rather treats the sentence "Mary believes that p" as more like something uttered in the course of telling a story than an assertion about a story. Story operators are employed when one is explicitly talking *about* a fiction, but when one is *telling* a story one is engaged in a quite different activity. Nothing approximating "According to the Perrault story, a

pumpkin turned into a coach" (or "Let's pretend that a pumpkin turned into a coach") is uttered in the course of *telling* Perrault's story. When one tells the story, one does not assert that a pumpkin turned into a coach; one does something more like pretending to assert it (Searle 1975; Lamarque 1981, 332). The analogous psychological noncognitivist fictionalist thinks that we do not (or need not) assert that Mary believes that *p*, but rather we do something more like pretending to assert it. The view counts as noncognitivist in virtue of the fact that it interprets an utterance of "Mary believes that *p*" as having assertoric force withheld. As with the cognitivist fictionalist interpretation, this removes ontological commitment to beliefs (etc.).

4. Unsuspecting Fictionalizing

None of these views in its hermeneutic form is committed to holding that ordinary speakers are consciously aware of what is claimed of their discourse. The cognitivist fictionalist need not maintain that ordinary speakers *think* "According to folk psychology. . . ." The noncognitivist fictionalist need not maintain that ordinary speakers *consider* their utterances to be make-believe assertions rather than real assertions. Rather, these are intended as charitable interpretations of their linguistic practices motivated by the ambition of avoiding the error theory. (Compare Harman and Thomson 1996, 4.) Of course, it remains to be seen whether any of these views is adequate. There may be strong reasons against interpreting belief claims as containing a tacit story operator or as make-believe assertions—reasons that outweigh the (supposed) advantages of avoiding error.

The revolutionary fictionalist, by contrast, seems committed to the fictive attitude being something of which the speaker may be aware. After all, the revolutionary fictionalist is recommending a *change* between the prerevolutionary erroneous utterances and the postrevolutionary trouble-free fictionalist utterances. If the fictionalist account were construed as a charitable interpretation that can be made of the postrevolutionary discourse (despite speaker ignorance of the fact), and there is no discernible difference between the prerevolutionary speaker and the postrevolutionary speaker, then there seems nothing to prevent that same charitable interpretation being made of the prerevolutionary discourse—in which case, revolutionary fictionalism would collapse into hermeneutic fictionalism.

However, saying that the difference between belief and make-believe is "something of which the speaker may be aware" is not to say that the

speaker is constantly aware of the difference. What typically distinguishes acts of make-believe, for example, is the fact that when push comes to shove they will be abandoned in favor of belief. We can be confident that Mrs. Lincoln's enjoyable engagement with the play ceased the moment John Wilkes Booth so rudely interrupted the evening's entertainment. But the fact that one *will* abandon one's make-believe (in such-and-such circumstances) is a dispositional property that may not have any phenomenological bite. Suppose a speaker goes around uttering sentence *p* in everyday life, but has the disposition to assert *not-p* if asked in an appropriately serious way—a way that takes into consideration skeptical arguments against *p*, etc. This may suffice for us to conclude that the speaker doesn't wholeheartedly believe that *p*; indeed, we might say that *p* is being treated as a kind of make-believe. That the speaker instantiates this disposition is something of which she *could* be made aware, but it's not something of which we are forced to say she *must* be aware. Therefore one might be engaged in an act of make-believe without being aware of the fact.

A different argument to much the same conclusion focuses on the fact that which speech act a speaker performs is not determined solely by speaker intentions; the conventions of the wider linguistic community also play a major role. Suppose a newcomer to our community is introduced to the word "slut," yet is taught it as if it's a perfectly descriptive term (denoting a woman with multiple sex partners, say) with no pejorative connotations. (This would not be teaching the person competence with the term.) The speaker goes forth and innocently uses the word whenever he needs to refer to a woman with multiple sex partners. When he employs the term on a given occasion, can we say that he succeeds in using it in a purely descriptive fashion, simply because that's his intention? I wouldn't think so. Does he, rather, use the term in a pejorative fashion, despite his intentions to the contrary? Now I am not so sure, but there is at least something to be said in favor of the positive answer. Those to whom he has talked will most naturally report "He said really insulting things," and can reasonably continue to describe matters in this way even when they realize the misunderstanding ("He said really insulting things, even though I now see that he didn't intend to"). Much the same thing can be said, I think, about acts of make-believe. A linguistic community may have a convention in place that decrees that when a particular phrase is

used (something like "Once upon a time. . .") an act of make-believe is entered into. But a token individual may be unaware of this convention (for whatever reason), and utter the phrase without intending to introduce an act of make-believe. Despite this lack of speaker intention, the surrounding conventions may dominate, ruling that what the speaker uttered after this phrase *is* make-believe. If this is correct, then we see again that one might be engaged in an act of make-believe without being aware of the fact; though, again, this is not to deny that any speaker may be brought to be aware of the fact.

5. The Benefits of Fictionalism

Let us now ask the obvious question: What would be the advantage of taking a fictive attitude toward folk psychology? If the theory of folk psychology, taken at face value, is false, then surely the honest thing to do is simply to declare it as such and eliminate it from our common practices? The fictionalist answer must be that folk psychology is *useful*, such that the cost of eliminating it is greater than the cost of taking some kind of fictive attitude toward it. (This is not to claim that the benefits of taking a fictive attitude toward folk psychology are as great as the benefits of believing it. But we are assuming here that the arguments for eliminative materialism have refuted the thesis that folk-psychological claims can be treated as true, if taken at face value as evincing ontological commitment to psychological entities.)

We must tread carefully hereabouts, since the claim that folk psychology has *not* pulled its weight in the practical sphere has been often taken as evidence in favor of ontological eliminativism. Paul Churchland says of the history of folk psychology that "the story is one of retreat, infertility, and decadence"—that an assessment of its explanatory success reveals "a very long period of stagnation and infertility" (1981, 74).[5] The psychological fictionalist embraces ontological eliminativism (and thus is presumably persuaded by arguments in its favor) and yet at the same time recommends maintaining folk psychology as a fiction on the grounds of its usefulness. There are several ways of navigating through this apparent tension.

First, the fictionalist may accept that folk psychology has not been sufficiently useful to warrant our thinking it true, but nevertheless allow that it has *some* practical uses—uses that provide its fictional adoption with a practical edge over linguistic eliminativism. After all, the kind of

failure to which Churchland draws attention is the failure of a scientific theory (regarding, e.g., explanatory success and "coherence and continuity with fertile and well-established theories in adjacent and overlapping domains" [1981, 73]), and he may be correct that such criteria are "the final measure of any hypothesis," such that failure suggests falsehood. But something may disappoint by this criterion yet still be very useful in other ways. Consider the works of Shakespeare.

Second, the fictionalist may have become convinced of the bankruptcy of folk psychology (taken at face value) on grounds other than its practical failure. He or she may accept all along that folk psychology has been and remains a jolly useful theory, but nevertheless have other reasons for thinking it false. This second strategy would be generally untenable only if one endorsed a kind of pragmatism about truth that seeks to exclude the very idea of the *false but useful*. While I have no problem with certain tenets of pragmatism, any version that is so gross as to countenance excluding this idea should have few adherents. (If I point a gun at you and demand that you utter "1 + 2 = 5," does this proposition—suddenly so acutely useful to you—become true [for you]?) Everybody sensible allows the possibility of false propositions whose utterance may be in some manner useful; and to this extent there is an accepted paradigm to which the fictionalist can appeal.[6]

6. Relativistic Fictionalism

With talk of usefulness inevitably comes the possibility of relativism, for what is useful to one group or one individual is not useful to another group or individual. I will discuss this matter with respect to hermeneutic and revolutionary fictionalism in turn.

Consider first hermeneutic fictionalism. Why would one prefer a fictionalist construal of the discourse over the more orthodox error theory which follows (we are assuming) if belief and desire claims are taken at face value? The fictionalist's answer is that there is some cost to allowing all utterances of the form "S believes that p" (say) to count as false, and so ordinary speakers shall instead be interpreted as engaging with a fiction. But suppose that there is some individual—call him "Paul"—for whom this is not true. Paul is unusually situated such that the falsity of all sentences of folk psychology suits him very well. Or we might imagine that Paul is part of a group of whom this is true. Or we might imagine a whole society of whom this is true—not *our* society, presumably, nor even

any actual society; but at least a possible society for whom ontological and linguistic eliminativism about folk psychology are no great burden.

It is not clear to what extent we must embrace relativism here: interpreting some speakers' utterances in a fictionalist manner but not other speakers'. What the hermeneutic fictionalist is doing is offering an interpretation of a linguistic *practice*, and if we are to interpret Paul's utterances as part of that same practice, then he may just have to put up with having the same "charitable" interpretation made of his utterances as everyone else, even though it may not suit him personally. In the same way, if some fool happens to think that the word "dog" denotes felines, the fact that it would suit her very well if it *did* denote felines (for then she would triumphantly be proven right) is insufficient to underwrite a relativism about reference such that from her lips "dog" means feline while for everyone else it means canine. Things may stand differently if there's a *group* who starts to use "dog" to denote felines. Then they would create a linguistic convention that might demand a special interpretation. (Think of surfers describing an impressive wave as "bad" or "wicked" or "sick.") Perhaps the same would be true of a group whose interests would not be served by rescuing folk-psychological discourse from error; perhaps their utterances of the form "S believes that *p*" (etc.) shouldn't receive the fictionalist interpretation.[7]

As for imaginary linguistic communities, I suspect that there just isn't a settled truth to the matter of whether we must interpret their utterances relative to what is most useful/charitable *for them*, or give the same interpretation of their utterances as we do of ours, which are interpreted relative to what is most useful/charitable *for us*. Even though there might exist a linguistic community at another possible world that would be better off having its own folk psychology sentences interpreted at face value (thus rendering them false), nevertheless *we* are interested in providing the best interpretation of *our* linguistic practices, and we are entitled to employ that interpretation when considering counterpart discourses at nearby possible worlds. If a fictionalist interpretation (of either a cognitivist or noncognitivist flavor) of actual folk-psychological discourse is warranted on pragmatic grounds, then we will likely impose it on imaginary folk-psychological discourses, even ones for which the cost-benefit analysis works out very differently from our own.

Consider now revolutionary fictionalism. Assuming that an error theory is true of actual folk-psychological discourse, why would one rec-

ommend some form of fictionalist response over the more orthodox lin-
guistic eliminativism? The fictionalist's answer is that there is some cost
to dropping folk-psychological discourse entirely from our lives, and so
ordinary speakers are advised to recoup some of those costs by engaging
with a fiction. Here a degree of relativism seems inevitable, for the revo-
lutionary fictionalist is essentially offering *advice*, and what is good
advice for one person or group may simply not be for another individual
or group, due to their being differently situated. Certainly we can at the
very least *imagine* whole populations for whom the fictionalist option is
suboptimal. Perhaps these are populations for whom linguistic elimina-
tivism regarding folk psychology is a negligible liability, or perhaps for
them the adoption of the fictionalist option brings some large hidden cost.
Indeed, it would not be surprising for the fictionalist to think of the fictive
attitude as a kind of interim arrangement: as warranted for us now (us for
whom folk psychology is an entrenched habit, but who have, let's say, just
come to accept ontological eliminativism about psychological entities),
but as something that might eventually be discarded as we leave folk psy-
chology entirely behind us. In other words, making a fiction of folk
psychology may be good advice relative to us now, but may be poor
advice relative to some future Brave New World.[8]

   Relativism regarding *individuals* is a more complicated affair, some
of the reasons for which have already been touched upon. Suppose Paul is
personally ready to embrace linguistic eliminativism about folk psychol-
ogy: no need for any spineless fictionalist nonsense for him! But let's say
that he is unusual; pretty much everyone else would find linguistic elimi-
nativism overly burdensome—more burdensome, at least, than keeping
folk psychology alive as a fiction. I see no problem in concluding that rev-
olutionary fictionalism is reasonable relative to most people in this
population but not reasonable relative to Paul.[9] Paul can choose simply
not to employ the terms of folk psychology, even though they are being
widely used around him. (Think of an atheist in a community of devout
theists.) This may be challenging for Paul in practical terms (what's he
going to say when asked "Do you believe that it will stop raining soon?"),
but perhaps he can find ways of getting by. One complication that this
reveals is that when one weighs the costs and benefits of fictionalism
versus eliminativism for an individual, what the other individuals are
doing in this respect is an important factor. Maybe eliminativism would

be optimal for Paul so long as a sufficient number of his fellows join him, but perhaps if he is surrounded entirely by fictionalists then the inconveniences of being constantly misunderstood speak in favor of his taking the same option as them. (In other words, perhaps the question of which is the optimal attitude for an individual to adopt exhibits some of the complexities of frequency-dependent Darwinian selection.)

More complications are revealed if we try to imagine the reverse: an individual for whom making a fiction of folk psychology is optimal, who is surrounded by a sea of linguistic eliminativists. The problem here is that it is not obvious that one can properly make a fiction of a discourse *by oneself*; a linguistic community is needed in order to create the necessary linguistic conventions. Imagine someone in our midst going round talking about phlogiston—pointing at flames and saying "Here is the phlogiston escaping." We would naturally take her to be making badly mistaken assertions, to be embracing a false theory. Could it be, though, that she is just pretending to assert these things, or that she is really making true claims with a tacit story operator ("According to phlogiston theory, here is the phlogiston escaping")? If she makes no effort to inform us what's going on, then it is not clear to me that the answer to either question is "Yes." As I suggested earlier, the conventions of the surrounding linguistic population may dominate the speaker's intentions and determine what speech act is performed. A person may intend to *pretend* to assert X, but if everyone takes her to be really asserting X then arguably she really does assert X. A person who performs a paradigmatically rude act, like spitting on the carpet, cannot defend himself by saying "I wasn't really being rude; I was just pretending to be." Or spare a thought for the poor genius who invented sarcasm: imagine him or her going round trying out the new tone of voice on people without explaining how it works.[10]

In sum, the revolutionary fictionalist's recommendation may be reasonable relative to some individuals and unreasonable relative to others, but in some cases the criteria that determine the matter may include that the individual is a member of a like-minded group of sufficient size and structure to underwrite a certain kind of linguistic convention.

7. Fictionalist Suicide

Many readers will by now be impatient to see a glaring worry with psychological fictionalism addressed: that the thesis is incoherent in that

it recommends that we adopt psychological states of which at the same time it doubts the existence. The articulation of the problem starts with ontological eliminativism, which rejects the existence of propositional attitudes like beliefs, desires, hopes, fears, dreads, wishes, and so forth. The cognitivist fictionalist, however, makes essential reference to *belief*: either (a) interpreting psychological claims (e.g., "Mary desires that *p*") as expressing *beliefs* about what is true according to folk psychology, or (b) recommending that we change our attitude toward such claims such that they come to express such beliefs. The noncognitivist fictionalist, by contrast, makes reference to *make-believe* rather than belief. But it seems plausible to maintain that make-believe is another category of folk psychology (see Currie 1990; Nichols and Stich 2003)—or at the very least that it is a kind of mental state of which we can make sense only within a framework that includes the propositional attitudes of folk psychology—in which case noncognitivist fictionalism also seems to be describing or recommending a kind of mental state whose existence it denies.

This problem of *fictionalist suicide* is reminiscent of a complaint that has been made against ontological psychological eliminativism in general: that it commits *cognitive suicide*. (See Cling 1989; Boghossian 1990; Devitt and Rey 1991; Reppert 1992.) There the criticism is that the ontological eliminativist denies the existence of beliefs, but in articulating and arguing for this thesis presumably puts forward propositions that express his/her *beliefs* and are recommended to the audience as items to be *believed*. In my opinion, the Churchlands responded successfully to this complaint long ago, with an especially effective parody involving the impossibility of anyone coherently denying the existence of vitalistic life force:

> The anti-vitalist says that there is no such thing as vital spirit. This claim is self-refuting; the speaker can expect to be taken seriously only if his claim cannot. For if the claim is true, then the speaker does not have vital spirit, and must be dead. But since dead men tell no tales, they do not tell anti-vitalist ones either. One cannot reason with dead men. (Patricia Churchland 1981, 100)

The argument is obviously question-begging. Someone who denies the existence of vitalistic life force at the same time denies that he (the person advocating the argument) must have vitalistic life force in order to present a meaningful argument. In the same way, the ontological eliminativist about beliefs not only denies their existence but also denies that what goes on when an argument is articulated, advocated, and accepted in

any way presupposes the existence of beliefs. What the articulation, advocacy, and acceptance of an argument *does* involve is, quite literally, hard to say. It is hard to say because what it really involves are complex neurological states about which our knowledge is patchy and for which ordinary language is lacking.

The challenge of fictionalist suicide, however, is not so easily overcome as the challenge of cognitive suicide. That ontological eliminativism is put forward as an item to be *believed* is not itself part of the content nor a presupposition of that theory. By contrast, and more worryingly, that we are able to *make-believe* that a fiction is true *is* an essential presupposition of certain forms of fictionalism. (Likewise, mutatis mutandis, with other forms of fictionalism.) This challenge is unique to psychological fictionalism; it doesn't arise for moral fictionalism, or modal fictionalism, or mathematical fictionalism, and so on. This is because the very statement of any fictionalist theory (such as "Make-believe that error-laden discourse X is true") involves reference to psychological entities, but not to moral, modal, or mathematical (etc.) entities.

One way someone might try to overcome this challenge is by denying that the kind of mental states required by fictionalism are among those rejected by the ontological eliminativist.[11] Rather than "make-believe," the key fictive attitude is sometimes called "acceptance"—a kind of embrace that falls short of belief (Van Fraassen 1980). The ontological eliminativist holds that the folk-psychological concept of *belief* indispensably involves a set of criteria that nothing satisfies. (William Alston [1996], for example, lists half a dozen plausible contenders for such criteria.) *Acceptance*, by contrast, might be taken to be something that requires some but not all of these criteria, such that (a) certain actual psychological states do satisfy the criteria for acceptance, and (b) acceptance is not a folk-psychological notion. Thus, one might maintain that acceptance is exactly what should replace the flawed folk-psychological notion of belief. The ontological eliminativist, accordingly, may have no objection to *acceptance*, and is thus free to recommend that we adopt this attitude, or to analyze extant states (erroneously taken to be belief) as being in fact states of acceptance.

The problem with this way of overcoming the challenge of fictionalist suicide is that it would achieve far too much. Someone who interprets those states which would ordinarily be considered *belief* to be instead

instances of *acceptance* presumably intends this construal to apply across the board. Thus, such an interpretation wouldn't be limited to the fictionalist's proposal that we accept (but not believe) folk-psychological claims, but seems committed to an infinitude of similar proposals: that we accept (but not believe) that gold has the atomic number 79, that we accept (but not believe) that Napoleon lost Waterloo, that we accept (but not believe) that the number 84 succeeds the number 83, and so on. My complaint is not that this is an incoherent (or even particularly implausible) perspective to adopt, but that the fictionalist proposal under consideration—pertaining to the attitude that should be adopted toward folk psychology—would lose identity as a distinct theoretical position.

Let me turn to another way that the psychological fictionalist might try to overcome the challenge.[12] Begin by recalling what I said a little earlier about the puzzle of cognitive suicide. Let's allow that the ontological eliminativist does articulate, advocate, and expect her audience to accept her theory, but she does not thereby undermine her theory since these phenomena (i.e., articulation, advocacy, and acceptance) are to be understood not as involving beliefs, but as involving neurological events for which we lack natural vocabulary. Thus, the ontological eliminativist may allow that when we would ordinarily take ourselves to have grounds for claiming "Mary believes that p" and "John believes that p," there is in all probability something similar going on in Mary's brain to what is going on in John's brain. And were we instead to take ourselves to have grounds for asserting "Mary desires that q," there will be yet another kind of neurological event occurring in Mary's brain. But (the eliminativist thinks) these neurological states do not deserve the labels "belief" and "desire" (and nor does anything else). (The eliminativist may be mistaken about this—some would prefer to identify these neurological states with the entities of folk psychology—but it is not my intention here to assess the eliminativist's case but rather to provisionally grant it.)[13]

An analogy may help. We are all, I take it, error theorists about witches. Yet it is possible that all and only those women who were accused of witchcraft did have some distinctive set of properties. Perhaps they were disempowered women who all threatened the patriarchal society in a certain manner.[14] Perhaps it was something more disjunctive and complicated than that. At the very least we know that there was one

property that all and only women accused of witchcraft instantiated: that of being a woman accused of witchcraft. Thus when people accused certain women of being witches, there was something distinctive going on with these women—there was some property or properties that they and only they instantiated—and yet it nevertheless remains perfectly reasonable to insist that there are no such things as witches. It would be madness to identify the property of *being accused of witchcraft* (or any property like *being a woman who threatens patriarchy in such-and-such a way*, etc.) with the property of *being a witch*. By the same token, when we employ folk psychology we "accuse" people of having beliefs and desires and so forth, and it may well be that there is something distinctive going on in their brains when we do this, but nevertheless (the ontological eliminativist thinks) it is a mistake to identify these complicated brain states with beliefs, desires, and so forth.

This solution to the challenge of cognitive suicide may encourage one to say something similar in response to the challenge of fictionalist suicide. The ontological eliminativist about psychological entities need not deny that when there is a paradigm instance (according to folk psychology) of someone's believing that *p*, there is something distinctive going on in the subject's brain: something that is different from what is going on in the brain of a person who is a paradigm instance (according to folk psychology) of someone's treating *p* as a make-believe. In other words, the ontological eliminativist need not hold that folk psychology is a *wild and whimsical* false theory utterly disconnected from what's going on in the world and in people's brains. Folk psychology probably imperfectly tracks real brain events; it's just that the conceptual apparati that folk psychology employs (beliefs, desires, etc.) are so removed from reality that it is best to deny that these entities exist. Perhaps another analogy would help. Phlogiston theory was not a wild and whimsical idea either; it was a solid scientific theory—but a false one. Paradigm instances of phlogiston escaping (as identified by the theory) really did map onto real phenomena: namely, instances of oxygen being consumed through combustion.

What then of psychological fictionalism? When there is a paradigm instance (according to folk psychology) of someone's believing that *p*, then let's just say that this person is in brain state B; and when there is a paradigm instance (according to folk psychology) of someone's make-

believing that *p*, then let's say that this person is in brain state M. Psychological fictionalism can, then, be translated into claims that are acceptable to an ontological eliminativist. Instead of recommending that we make-believe rather than believe propositions of the form "Mary believes that *p*," the fictionalist recommends that we adopt brain state M rather than brain state B toward such propositions. Instead of analyzing propositions of the form "Mary believes that *p*" as "According to folk psychology, Mary believes that *p*," and then claiming that the latter may be truly believed, the fictionalist can say that instead of taking brain state B toward the proposition "Mary believes that *p*," we should take brain state B toward the proposition "According to folk psychology, Mary believes that *p*."[15]

The fact that psychological fictionalist proposals can in principle be imperfectly redescribed as claims about brain states does mean that something "in the spirit" of psychological fictionalism may be true or reasonable. However, inasmuch as the thesis of fictionalism is essentially characterized with reference to folk-psychological entities, then if the brain states in question cannot be identified with the entities of folk psychology (as the ontological eliminativist insists they cannot be), then psychological fictionalism simply cannot be literally correct. For example, if one denies the existence of make-believe, then one simply cannot propose that a person should take an attitude literally of make-believe towards X—not, at least, if the recommendation is supposed to be one that could be complied with. If one denies the existence of belief, then one cannot propose that a discourse consists literally of assertions containing a tacit story operator—not if to assert something is to express a belief in that thing. Strictly speaking, then, one might continue to insist that psychological fictionalism, as it is usually articulated, does indeed commit suicide, even if some other theory in a similar "spirit" remains viable.

But despite its suicidal tendencies, psychological fictionalism can be in some measure resurrected. As we have seen, the ontological eliminativist may allow that the analyses and proposals of fictionalism can be taken to roughly concern real brain states. Lacking a proper vocabulary for discussing these brain states with the appropriate specificity, however, one seems justified in using the best shorthand available for denoting them—namely, the terms of folk psychology. The fictionalist is not proposing that we identify, say, *brain state M* with *acts of make-believe* (for that

would be to abandon ontological eliminativism), but rather that we use the erroneous terms as a kind of imperfect proxy for complicated phenomena that we might otherwise struggle to describe. In a similar way, an anthropologist studying a foreign culture may find it convenient simply to speak of those actions and objects that are "tapu" (something in which he disbelieves), rather than laboriously referring to "what the people in this tribe tend to classify as tapu."

When someone in this manner says something that he himself judges to be a false, for the sake of convenience, is this a kind of fictive act? Not necessarily. If I threaten to shoot someone unless he utters the sentence "1 + 2 = 5," then his prudent decision to comply does not really count as a fictive act. Such an act of linguistic convenience wouldn't involve the speaker *immersing* himself in any fiction; it doesn't involve him allowing falsehoods a role in deliberations. Moreover, we need not classify his utterance as one that we would ordinarily think of as ontologically committing but which is not ontologically committing, for I think we are all pretty clear that what a person says with a gun to his head needn't be anything he really means.

Similarly, when someone wishes to recommend adopting brain state B (or M, etc.), but lacks the vocabulary for doing so and thus speaks in terms of "belief" (or "make-believe," etc.) this act of linguistic convenience is not itself a fictive performance—not, at least, in the sense in which we're interested. Such a speaker is not herself following any fictionalist advice; she isn't conforming to any recommendation to make-believe that p, or to believe that according to F, p. Thus the psychological fictionalist may be justified in articulating the theory in exactly the way it is usually articulated: using terms like "belief," "make-believe," "acceptance," and "assertion." This will not be to speak literally; the fictionalist is adopting a language of convenience when she uses these terms to present her favored theory. But in using this familiar language the fictionalist is not presupposing the accuracy of any of the conceptual machinery of folk psychology. There are literal recommendations and analyses standing behind these specious terms, just not a convenient language with which to express them.[16]

Richard Joyce

Victoria University of Wellington

Notes

1. One can be an eliminative materialist about different categories of mental entity. Daniel Dennett, for example, denies the existence not of propositional attitudes but qualia (Dennett 1988). Nevertheless, in this paper I focus exclusively on the case against propositional attitudes, and even there mention only beliefs and desires. Speaking of Dennett, it should be noted that the common assumption that his "intentional stance" is a kind of fictionalism is mistaken (Dennett 1987). He is adamant that his psychological instrumentalism allows for the real existence of beliefs and desires, but as "abstracta" rather than neurological events. "Some instrumentalists have endorsed *fictionalism*," he writes, and immediately declares that his instrumentalism is of a different stripe (1987, 72).

2. The Churchlands' talk of "theory replacement" is indeterminate between (i) replacing one ontological theory with another, and (ii) replacing one linguistic practice with another. The possibility of fictionalism forces one to make this distinction.

3. It is *possible* to be a hermeneutic fictionalist about Xs without being an ontological eliminativist about Xs, but such a view is unusual and hard to motivate. Such a theory holds that Xs really exist, but that whenever we speak of Xs we do not make assertions that commit us to the existence of Xs. An attempt to advocate this view (or one like it) is made by Christopher Jay, in his PhD thesis *Realistic Fictionalism* (University College London, submitted 2011).

4. Another view that might be categorized similarly to Ryle's, given our current taxonomic concerns, is that of Wilfrid Sellars (Paul Churchland's Ph.D. supervisor). Sellars (1956) argues that an utterance of "Mary believes that *p*" functions not to state an empirical fact but rather to perform a normative act of justification. (Of course, Sellars's view has many subtleties and is open to interpretation.)

5. Critics of Churchland on this point include Horgan and Woodward (1985) and Lahav (1992).

6. See my critical discussion of Hilary Putnam's thinking on this matter (Putnam 1971, 68ff) in Joyce (2001, 189–90).

7. The analogy may be importantly misleading. The speaker who misuses "dog" is making a *semantic* error. The question of whether "S believes that *p*" should be interpreted as "According to folk psychology, S believes that *p*" is also, I take it, a matter of the *semantics* of the predicate ". . . believes that. . . ." However, the question of whether a speaker is, when uttering "S believes that *p*," *asserting it or pretending to assert it* is a matter of the *pragmatics* of the language. Hence the "linguistic conventions" of which I airily speak may be quite different in the different cases, such that the point about an individual's intentions being dominated by the surrounding linguistic population may hold in one case but not the other. I don't have the space to pursue and straighten out the nuances here.

8. The idea of the fictive attitude (toward a variety of things) as an interim stance that lies between naive belief and enlightened disbelief, is inherent in the work of Hans Vaihinger (1949).

9. Since revolutionary fictionalism is a practical recommendation, the question of relativism must be expressed as "what is *reasonable*, relative to X," rather than "what is *true*, relative to X."

10. Of course, I don't seriously think that anything remotely like this occurred when the conventions of sarcasm emerged.

11. I thank Stuart Brock for pointing out this possibility to me.

12. Here I develop thoughts presented by Meg Wallace, in an unpublished manuscript.

13. Quine writes:

> Is physicalism a repudiation of mental objects after all, or a theory of them? Does it repudiate the mental state of pain or anger in favor of its physical concomitant, or does it identify the mental state with a state of the physical organism (and so a state of the physical organism with the mental state)? (Quine 1960, 265).

He thinks the question that forces the choice is misguided.

14. I've used this example before in Joyce (2001, 96; 2011, 531–33).

15. Wallace writes:

> [I]t is not that the Mental Fictionalist thinks that absolutely *nothing* is happening in the world when someone claims to be pretending she is Superman or that she is drinking tea with her imaginary friend. There *is* cognitive activity when these things are seemingly going on. It is just that our ordinary, everyday pretense-sincere avowal talk is wildly disparate from the cognitive activity that *is* going on—so much so, that terms such as "pretense," "make-believe," and "sincere avowal" fail to pick out any activity or process that's actually in the world. However, the advantage of Mental Fictionalism is that it allows one to still *talk as if* these terms do pick out something, even though this "talk as if" will ultimately (and strictly speaking) get cashed out in terms of some sort of complicated cognitive activity that is unsuitable as a legitimate element of Folk Psychology. (unpublished)

16. I am indebted to Stuart Brock and an anonymous referee for useful feedback.

References

Alston, William 1996. "Belief, Acceptance, and Religious Faith," in J. Jordan and D. Howard-Snyder, eds., *Faith, Freedom, and Rationality*, Rowman and Littlefield, 3–27.

Boghossian, Paul 1990. "The Status of Content," *Philosophical Review* 99: 157–84.

Churchland, Patricia 1981. "Is Determinism Self-Refuting?" *Mind* 90: 99–101.

———. 1986 *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, Cambridge, MA: MIT Press.

Churchland, Paul 1981. "Eliminative Materialism and the Propositional Attitudes," *Journal of Philosophy* 78: 67–90.

———. 1979. *Scientific Realism and the Plasticity of Mind*, New York: Cambridge University Press.

Cling, A. 1989. "Eliminative Materialism and Self-Referential Inconsistency," *Philosophical Studies* 56: 53–75.

Currie, Gregory 1990. *The Nature of Fiction*, Cambridge: Cambridge University Press.

Dennett, Daniel 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.

———. 1988 "Quining Qualia," in A. Marcel and E. Bisiach, eds., *Consciousness and Contemporary Science*, Oxford: Oxford University Press, 42–77.

Devitt, Michael and Georges Rey 1991. "Transcending Transcendentalism," *Pacific Philosophical Quarterly* 72: 87–100.

Harman, Gilbert and Judith Thomson 1996. *Moral Relativism and Moral Objectivity*, Cambridge, MA: Blackwell.

Horgan, Terence and James Woodward 1985. "Folk Psychology Is Here to Stay," *Philosophical Review* 94: 197–226.

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

———. 2011. "The Error in 'The Error in the Error Theory,'" *Australasian Journal of Philosophy* 89: 519–34.

Lahav, R. 1992. "The Amazing Predictive Power of Folk Psychology," *Australasian Journal of Philosophy* 70: 99–105.

Lamarque, P. 1981. "How Can We Fear and Pity Fictions?" *British Journal of Aesthetics* 21: 291–304.

Lewis, David 1978. "Truth in Fiction," *American Philosophical Quarterly* 15: 37–46.

Nichols, Shaun and Stephen Stich 2003. *Mindreading*, New York: Oxford University Press.

Putnam, Hilary 1971. *Philosophy of Logic*, New York: Harper and Row.

Quine, W.V. 1960. *Word and Object*, Cambridge, MA: MIT Press.

———. 1961. *From a Logical Point of View*, Cambridge, MA: Harvard University Press.

Ramsey, W., S. Stich, and J. Garon 1990. "Connectionism, Eliminativism, and the Future of Folk Psychology," in J. Tomberlin, ed., *Philosophical Perspectives: Action Theory and Philosophy of Mind*, Atascadero, CA: Ridgeview, 499–533.

Reppert, V. 1992. "Eliminative Materialism, Cognitive Suicide, and Begging the Question," *Metaphilosophy* 23: 378–92.

Rorty, Richard 1970. "In Defense of Eliminative Materialism," *Review of Metaphysics* 24: 112–21.

Ryle, Gilbert 1949. *The Concept of Mind,* London: Hutchinson.

Searle, J.R. 1975. "The Logical Status of Fictional Discourse," *New Literary History* 6: 319–32.

Sellars, Wilfrid 1956. "Empiricism and the Philosophy Of Mind," in H. Feigl and M. Scriven, eds., *Minnesota Studies in the Philosophy of Science, Volume I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Minnesota: University of Minnesota Press, 253–329.

Stich, Stephen 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief*, Cambridge, MA: MIT Press.

Taylor, K. 1994. "How Not to Refute Eliminative Materialism," *Philosophical Psychology* 7: 101–25.

Vaihinger, Hans 1949. *The Philosophy of 'As If '*, C.K. Ogden, trans., London: Routledge and Kegan Paul.

Van Fraassen, Bas 1980. *The Scientific Image*, Oxford: Oxford University Press.

Wallace, Meg. Unpublished manuscript. "Mental fictionalism."