

# Moral Fictionalism

Daniel Nolan (Syracuse University)  
Greg Restall (University of Melbourne)  
Caroline West (University of Sydney)

*Draft only.* Please do not cite without permission.

---

## Introduction

What would morality have to be like in order to answer to our everyday moral concepts? What are we committed to when we make moral claims such as “female infibulation is wrong”; or “we ought give money to famine relief”; or “we have a duty to not to harm others”, and when we go on to argue about these sorts of claims? It has seemed to many—and it seems plausible to us—that when we assert and argue about things such as these we presuppose at least the following.

*First*, that there is a realm of moral facts (in a non-technical sense of ‘facts’); facts in virtue of which claims like “female infibulation is wrong” are true and claims like “acts of senseless cruelty are right” are false. When we argue about whether female infibulation is wrong, for example, we seem to presuppose that that there is a fact of the matter about the rightness or wrongness of female infibulation about which we can form beliefs and which makes at most one of the disagreeing parties *right*. Furthermore, we suppose that the rightness or wrongness of female infibulation is something that we can discover. The assumption that there is a discoverable fact about the rightness or wrongness of female infibulation, for example, which our moral beliefs can get right or wrong, explains the purpose and nature of ordinary moral debate.

*Second*, we take these facts to be objective facts—the rightness or wrongness of female infibulation, for example, doesn’t depend on what we happen to think or feel. Bullfighting or torturing innocent children would still be wrong, even if we all thought or felt it to be right.

*Third*, many take it that these objective facts would have to possess prescriptive and motivational force in order to answer to our commonsense moral conceptions. If someone believes that they ought to give money to famine relief, to use a familiar example, then *ceteris paribus* they are inclined to give money to famine relief. Many take there to be something very strange and puzzling about someone who agrees that they ought to give money to famine relief, and yet insists

that they have no reason or inclination to do so. Of course—and this is the point of the *ceteris paribus* clause—they may not act on their reason or inclination: they may be weak willed, they may have false factual beliefs, there may be some other consideration which overrides this inclination. But, absent these sorts of countervailing forces, to have a genuine moral belief that one ought to give to famine relief is to have a *prima facie* motivation to give money to famine relief (or, at least, to have such a motivation insofar as one is rational). In sum, to borrow a phrase from John Mackie, our ordinary moral judgements seem to involve an implicit claim to the “objective prescriptivity” of moral facts.

These presuppositions lead to a well-known problem.<sup>1</sup> If this analysis is right, moral facts or properties would have to be very “queer” sorts of facts or properties indeed, utterly unlike any other features of the world with which we are familiar. It seems that moral facts or properties, if they were to be the sorts of objectively prescriptive things our moral practice and concepts seem to presuppose, might have to be strange *sui generis* entities, ontologically independent of any natural properties (like Plato’s forms), the apprehension of which necessarily impacts upon our desires. This metaphysical problem brings with it a corresponding epistemological one: how could we ever come to have knowledge of these queer non-natural, objectively prescriptive facts? Presumably, it will not be in the usual way in which we discover the existence of natural properties: by looking through a microscope or bumping into them.

The problem, then, is that our moral talk seems realist and cognitivist in nature—it presupposes that there are moral truths, and our moral judgements express beliefs about these truths—but the ontology this talk seems to presuppose is strange, non-natural and Moorean. We have good reason to reject it. So we seem to face a dilemma—either embrace an alien ontology, which we have good reason to reject or become eliminativists about morality and agree that moral claims are all false, since there are no queer entities of the sort such claims presuppose. As a consequence, we ought to abandon false talk of rightness and wrongness, goodness and badness, duties and obligations, justice and injustice, altogether.

---

<sup>1</sup> See Mackie (1977) ch. 1, for the classic statement of the problem.

This eliminativist price has seemed to many too great a price to pay. Moral discourse is extraordinarily useful. Morality plays a very important social role in coordinating attitudes and in regulating interpersonal relations. Giving up moral talk would force large-scale changes to the way we talk, think and feel. These changes would be very difficult to make (witness many a committed nihilist who cannot seem to refrain from using moral language in discussions of female infibulation, domestic violence and the like).

We have, then, incentive for finding some way in which to retain our realist moral discourse without its accompanying undesirable commitments. Can this be done? The non-cognitivist claims that it can; and the desire to vindicate our realist-seeming moral talk, without thereby being committed to the spooky Moorean ontology, is a large part of what motivates non-cognitivist approaches.<sup>2</sup> But the non-cognitivist attempts to do so by denying that our moral talk is really as it seems: namely, realist talk. Non-cognitivists either have to deny very plausible things about what we take ourselves to mean when we speak morally or they are offering us a substitute for our current moral talk which merely changes the subject.<sup>3</sup>

But there is an alternative to such approaches that has been largely overlooked<sup>4</sup>; and we think that it is superior to many of its more

---

<sup>2</sup> Of course, another common anti-realist strategy, taken by non-cognitivists or emotivists such as A.J. Ayer, is to deny that realism is any part of the meaning of ordinary moral terms at all. Moral statements or propositions are not even superficially truth-apt, for they purport to express attitudes rather than to offer descriptions or reports.

<sup>3</sup> Another approach, this time a realist one, is to accept that ordinary moral discourse is as it seems, viz. genuinely cognitivist in nature; but to deny that this discourse presupposes the problematic non-natural ontology. Naturalistic moral realists (such as Sturgeon [1985]; Railton [1986]; Brink [1989]; Smith [1994]) claim that the truth-makers for moral claims are scientifically-respectable, ordinary natural properties or features that can be discovered by empirical observation or inferred from observational information—although they differ over the details of which natural properties are identical with the moral ones, and over the nature of the connection between these properties or beliefs about them, on the one hand, and motivation and action on the other. Naturalistic forms of moral realism face a number of challenges of their own (see Smith [2000]).

<sup>4</sup> While a number of philosophers—including Critias (see Bury, 1936, esp. p.31), Plato (1993, esp. p. 119), Mandeville (1924, esp. 28-44), Bentham (see Ogden, 1932), Nietzsche (1968, esp. Book Two, Part II), Mackie (1977), and perhaps even Quine at one stage (1969, esp. p. 146)—make remarks that suggest that their view of ethics, or a close relative of their view, can naturally be construed as a fictionalist approach of one form or another, very few philosophers have explicitly endorsed a moral fictionalist position, especially if one sets aside the version of fictionalism that merely recommends deceit about moral matters. The neo-Kantian Hans Vaihinger (1935) explicitly proposes a moral fictionalism, suggesting a possible reading of Kant along moral fictionalist lines.

familiar rivals. We might both accept that our moral talk is as it appears to be (with the realist and against the non-cognitivist) and that it is false (with the eliminativist), but nevertheless think that such realist moral talk ought be to retained because it is useful.<sup>5</sup> The result is *fictionalism about moral discourse*, or more succinctly, *moral fictionalism*.

In this paper we outline fictionalism in general, giving a general explanation of how it works, along with some of its advantages and disadvantages. Against this backdrop, we explore moral fictionalism and suggest why it may be preferable to some better-known anti-realist and quasi-realist approaches.

We begin in Section 1 with an introduction to fictionalist approaches in general, setting the scene for *moral* fictionalism by discussing some well-known contemporary fictionalist approaches to discourse in other, non-moral domains; and clarifying how a fictionalist deals with the issue of ‘assertibility’ (just when is it appropriate to utter untruths such as ‘x is good?’). In Section 2, we turn to *defining* moral fictionalism, illustrating the position with some examples. In Section 3 we compare the fictionalist approach to moral discourse with more orthodox anti-realist and quasi-realist positions, and explain why the fictionalist approach may be preferable. Notwithstanding its advantages over near rivals, moral fictionalism faces its own special challenges. In Section 4, we discuss these challenges and sketch some possible fictionalist lines of reply.

## 1. Introducing Fictionalism

The simplest fictionalist approach to a discourse is a matter of taking certain claims in that discourse to be literally false, but nevertheless worth uttering in certain contexts, since the pretence that such claims are true is worthwhile for theoretical purposes. Fictionalism should

---

<sup>5</sup> There is one reading of Mackie according to which Mackie himself is advocating this sort of fictionalist position (Garner [1993] and Jackson [1994] suggest this interpretation). Many have puzzled about what is going on when, having made the case for thinking that all moral claims are false, Mackie then goes on to develop a "practical morality". One way of explaining this apparent inconsistency is to view Mackie as a fictionalist: we should (pragmatically) continue to use the old, infected moral vocabulary in order to fool the vulgar into behaving. An alternative, non-fictionalist explanation of what is going on is that Mackie thought that, in the absence of morality, we need some set of conventions which play the coordinating role that talk of morality currently plays, and in virtue of playing a similar role these conventions can be moral norms in a loose sense (although strictly speaking perhaps they should have been called moral\* norms).

thus be distinguished from more general nihilist or error-theoretical approaches, since often someone proposing an error theory in a certain area will recommend that we cease using the claims in question. When it was discovered that there was no phlogiston, scientists recommended no longer using phlogiston theory. On the other hand a fictionalist about phlogiston would recommend that we pretend that the relevant claims about phlogiston are true, for some theoretical purpose. As a borderline case, note that there are those who act like fictionalists insofar as they employ theories that they do not believe are true, but who merely hold that they do not know whether or not the relevant claims are true. (A case is van Fraassen's instrumentalist attitude to unobservable physical objects). We shall extend the rubric of "fictionalist" to cover such cases, though nothing hangs on this. For simplicity, though, we shall normally assume the fictionalist treats the claims treated fictionally to be literally false.

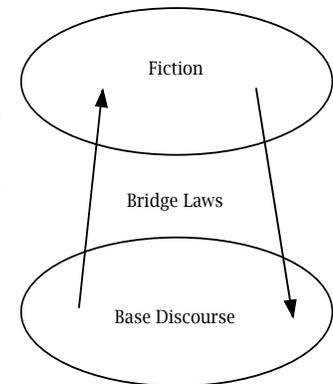
Paradigm instances of fiction have always attracted fictionalist approaches: even though we know stories of hobbits are false, we find it useful to say things that appear to commit us to them nonetheless. The use might be as simple as entertainment or literary criticism, or there may sometimes be more didactic purposes in saying things believed to be only fictions: cautionary tales, Aesop's fables, stories which provide insight into the human condition, and so on. We are fictionalists about the claims of Orwell's *Animal Farm*, since we think that claims about Boxer's loyalty or Napoleon's unspoken plans are worth making for a variety of useful purposes: as entertainment, to discuss our appreciation of the book, or to warn against Stalinist totalitarianism. We acknowledge that there really was no such place as the farm taken over by talking farm-animals. A starting point for understanding the fictionalist attitude to theories treated fictionally is to take it that they treat those claims as they treat the claims in paradigm fictions.

Fictionalism is more interesting, and more controversial, in areas where some are inclined to take what is said at face value. Perhaps the best-known fictionalist position in recent philosophy is the position taken by Hartry Field with regard to mathematics. In *Science Without Numbers and Realism, Mathematics and Modality*, Field takes the view that the arithmetical claims we make which would rely for their truth on the existence of numbers, or functions, or sets, or features of such things, are all *false*. Literally speaking, for Field, there are no numbers. Literally, there is no prime number between nine and twelve, for

example, and no commutative addition function, for there are no numbers or functions at all. This does not mean that we should cancel mathematics classes. The theory of mathematics, which the realist takes to be unvarnished truth is, for Field, a fiction. It is a convenient fiction, rather than a misleading one better forgotten, because we can go from information about the world and the objects in it to conclusions about the same subject matter, taking a detour through the fiction.

The classic terminology for this structure is as follows: the unproblematic literally interpreted theory is the *base discourse*. The *fiction* is a theory we take to be false, but useful nonetheless. Then *bridge principles* link the base discourse to the fiction. Typically, bridge principles are conditionals or biconditionals like this:

*x* is 5 metres long if and only if *according to the fiction x*  
has a *length* of 5 metres.



The left hand side of the biconditional is a simple predication of a property to *x*. The right hand side wears new ontology on its sleeve: it states that there is a thing called a *length*, which is 5 metres, and *x* has that *length*. The fiction is then a theory about *lengths*.

Care must be taken in adding the fictional discourse to the base discourse. If the base discourse is *rich* enough, you do not simply *add* the fiction with the bridge laws to the base discourse, for the addition may be inconsistent. If fictionalism can be *stated* in the language, then the base discourse will say that the claims in the base discourse are *false*.<sup>6</sup> These are claims of the base discourse. So adding the fiction will result in inconsistency. So, we must add the fiction and the bridge laws more subtly. For a fiction that merely posits new *ontology* (such as Field's fictionalism about numbers) we *protect* the base discourse by restricting the quantifiers of the base discourse with a new predicate (say *E* for existence) and we then add the bridge laws and the fiction.<sup>7</sup> If the

---

<sup>6</sup> Or, in the case of a fictionalism agnostic about the literal truth of the fiction, it will state things that the base discourse takes to be out of bounds for some reason.

<sup>7</sup> This strategy does not *ensure* that the result is consistent, of course. The fiction might make a claim directly in contradiction to the base discourse (think of fictions which make false claims about real people) or the bridge laws might lead to contradictions. For this case, think of a fictionalism about space according to which space is Euclidean. If there is no space (no points, no lines), then nothing in the

fiction posits new *properties* of the objects of the base discourse, we must add the new discourse with some other technique shielding the claims from conflict.<sup>8</sup> Whatever strategy we use to make this process precise, we will use the word “combining” for the *protected* addition of the fiction and the bridge to the base discourse.

For Field, a fictionalist strategy is useful because we may be assured that this detour through a fiction will leave no admixture of falsehood when we return to the base discourse. Starting from true premises about the world, and proceeding with mathematical reasoning, we can draw conclusions about the world and be guaranteed valid inferences. If the premises are true, then necessarily, so is the conclusion. Field offers us an illustration of how this might be done in *Science Without Numbers* by showing how a mathematical theory can be a *logically conservative extension* of a physical theory. Given a language capable of expressing various claims about physical objects, a language for making mathematical claims as well as some mathematical axioms and rules, and some “bridge principles” about how to go from physical claims to mathematical claims and vice versa, it can be shown that if a statement in the physical language is a logical consequence of a set of statements in the physical language together with the bridge laws and mathematical axioms, then it is also a logical consequence of the “physical” premises alone. In an area where this has been proved, we have a guarantee that treating the mathematical theory as if it were true, we shall not be led astray about the physical domain.

Field also has an answer for why we should avail ourselves of this conservative extension. Even in an area where in principle our conclusions are logical consequences of non-mathematical statements, showing this is incredibly cumbersome and difficult. (Field’s *Science*

---

fiction contradicts the base discourse. However, the addition of bridge laws (adding, say, “ $x$  and  $y$  and  $z$  are collinear if and only if an uninterrupted photon would pass through  $x$ ,  $y$  and  $z$ ”) could well make the theory inconsistent.

<sup>8</sup> There will be no problem if the predicates used in the fiction are not present in the language of the base discourse. However, in the case at issue for us, this may not be available as a strategy. We may want the base discourse to say truly “there are no lengths” or “there are no obligations.” To make the combined theory consistent, we need regimentation of the language to shield base claims from claims in the fiction. The simplest is to *tag* fictional predicates, using them as truly independent predicates. So, the combined theory says that there are no lengths, but that there are lengths\*. Then when it comes to interpreting the unregimented discourse, we must have a policy for determining which uses of predicates deserve asterisks and which do not. Typically this is marked by words such as “literally.” You say “It has a length of 5 metres. But of course, there are *literally* no lengths” which we understand as “It has a *length\** of 5 metres. There are no *lengths*”. This is consistent.

*Without Numbers* provides convincing cases of this.) Using mathematics is vital to *doing* science, or indeed dealing with our world in countless ways: if Field is right, not because we could not in principle dispense with it, but because it is the most convenient shortcut we have. Field's programme is still in its early stages, and he has only shown that mathematics can be treated as a dispensable add-on in limited areas of physics. Still, Field's approach is an example of a view which both explains one way in which we can be confident that employing a fiction for a serious theoretical purpose will not lead us astray, and why even so we should employ the fiction instead of dispensing with it.

If a fiction is a logically conservative extension of the truth, we have a *guarantee* of safety. But sometimes it may be desirable to employ a theoretical fiction that we have not proved is a logically conservative extension of the theory we believe, or even to employ one that we take to give us conclusions which cannot be deduced from our starting commitments alone. A well-known example of why one might take the risk using a theoretical fiction is provided by van Fraassen's attitude to unobservable physical objects.<sup>9</sup> Van Fraassen thinks we should be agnostic about the existence of such objects, even when the best physics, chemistry, biology or whatever say that such things exist. Instead of genuinely believing theories that involve such objects, van Fraassen recommends that we only treat these theories as if they were true for certain purposes, like prediction of the phenomena we *can* observe. In our terminology, van Fraassen is a fictionalist about unobservable objects. In his terms, we are to "accept" the claims about unobservables in our good scientific theories, but should withhold belief from these claims.

Like Field, van Fraassen can recognise a division into the "base discourse" which consists of statements we should take literally (those about observable objects), "bridge laws" for going between such claims (e.g. reports of scientific observations) and claims about unobservable objects, and theoretical claims solely about unobservable objects—which we will call the "fiction." The boundaries for these groupings might be somewhat vague, as van Fraassen admits<sup>10</sup>—but however they are drawn, the predictive power of modern science allows us to go from information about observables, to predictions about other

---

<sup>9</sup> van Fraassen (1980)

<sup>10</sup> See van Fraassen's discussion of the vagueness of the observable/unobservable distinction in van Fraassen (1980), pp. 15–19.

observable factors, via hypotheses about unobservables. These predictions, if they are to be interesting, must go beyond what logic alone can tell us given our original information: science can tell us about the future, or places we have never been, or the outcomes of situations which have never been created before. So adding our theory to observations can give us new information about what is available to be observed. A logically conservative extension of our observations, on the other hand, would not yield new information. So van Fraassen does not require that we can never draw new conclusions about observables from our previous stock of information about observables plus the “accepted” theory about unobservables.

While van Fraassen would not want our theory of unobservables to be a logically conservative extension of our direct evidence of observables, he does, of course, want the claims “accepted” about unobservables to be truth-preserving in a weaker sense: we would not want this theory to allow us to go from truths about observables, via the bridge principles and theory of unobservables, to conclusions about observables which were in fact false. Hinckfuss, in the context of defending fictionalism about space-time, calls an extension with the property of not in fact leading from truth to falsity a *physically* conservative extension.<sup>11</sup> For greater generality, we shall call this property the property of being a “truth-value preserving extension.” Indeed, while risking error is unavoidable, it seems to be an ideal at least to not be led into *new* errors by self-consciously employing a convenient fiction.

It might be asked what assurance we have of such a truth-value conserving extension, where a proof that an extension is logically conservative is unavailable or undesirable. A powerful reply of van Fraassen’s is a *tu quoque*.<sup>12</sup> What guarantee does a realist have that their commitments to unobservable objects will not lead them to errors about observable ones? The answer is likely to be a piecemeal one, in scientific terms—why we have good reason to believe that there are atoms and they behave in certain ways, and why, if there are and they behave in those ways, we should expect the predicted behaviour of observable objects. Van Fraassen can piggyback on the standard explanation: he can tell us what reasons we have to *accept* claims about atoms and their behaviour (the reasons the realist gives for belief), and

---

<sup>11</sup> Hinckfuss (1996)

<sup>12</sup> van Fraassen (1980), p. 36

good reasons to accept these claims are good reasons to suppose that they will not lead to falsehoods about observables. Predictive success over a range of cases will indicate that accepting a certain theory will not lead to false predictions, just as predictive failure will be evidence that the theory leads us into falsehood. So, justification for taking a theory to be truth-value conserving need not take the form of Field's admittedly conclusive route.

Of course, a fiction might not even be truth-value conserving, and yet *still* be useful. Conservativeness is undoubtedly a virtue for a fiction but it is no more essential for a fiction than truth is for a realistically interpreted theory, or empirical adequacy for an instrumentally interpreted theory. Newtonian mechanics is useful despite the inaccuracies of its predictions about the very small, the very large or the very fast. In the same way, a fiction might fail to be conservative while still retaining its uses. All we would require is that it does not lead from truth to falsehood *to a degree that matters*. Conservativeness is a virtue for fictions, yet it need not trump all others. Simplicity, fruitfulness, and other familiar virtues apply just as well for theories treated as fictions as those taken to be true.

Given the introduction of van Fraassen's instrumentalism as a kind of fiction, we can take this opportunity to clarify our commitments with respect to *theories* such as the fiction and the bridge laws. While we will typically talk of theories as a collection of sentences in some language, there is no reason to think of theories only in this way. If you agree, with van Fraassen, that theories are better or more realistically presented in terms of *models* then our fictionalism works for you too. There is no requirement that either the fiction or the bridge laws be presented as a set of sentences. Perhaps they are presented as models. No matter—all we need is some way of *grafting* the model or models determining the fiction onto the model or models used to determine the base discourse. The technique this grafting uses will be similar to the process we have called the "protected addition" of the fiction onto the base discourse. Typically, the mode of addition of models will be simple. We simply add the objects in the model of the fiction to that of the base discourse, and interpret the predicates of our language as our fiction demands. Nothing of interest for our purpose hangs on the presentation of theories.

Before we explain moral fictionalism in any detail, it is useful to clarify how fictionalists can deal with a further issue: assertibility.

### ***Assertibility***

What is involved when a proponent of a moral fiction makes a claim such as “*x* is good”? As fictionalists we take it that sometimes such assertions are warranted, despite being false when taken at face value. The situation with moral fiction is analogous to the assertion “Bilbo Baggins is short” or “there is a prime number between 2 and 5.” We want to say that some such assertions are reasonable and that others are not. Just when is it appropriate to say such falsehoods? A fictionalist can give a number of different answers to this question—we will examine the terrain without committing ourselves to any particular account of the reasonableness of “fictional assertions.”

The first bifurcation is between analyses that take reasonableness to be of a piece with assertibility and those that do not. If reasonableness is *not* analogous to assertibility, then there are at least two options. One is to take the norms to be instrumental norms of *deception*. Moral claims could be used with an intent to deceive and whenever it is appropriate to deceive an audience into thinking that they ought do *x*, it is appropriate to say “you ought do *x*.” Another option is to follow the non-cognitivist and express reasonableness conditions in terms something other than belief. It is appropriate for me to say “you ought do *x*” when I want you to do *x*.

These accounts are available to the fictionalist, but they do not utilise the resources available as a result of adopting fictionalism. One of fictionalism’s virtues is the smooth treatment of discourse: moral claims are as truth-apt as other claims. It also allows for a uniform treatment of appropriateness conditions, with close connection to assertibility for the literally true. If we take appropriateness for “fictional claims” to be like the norms of assertibility, again we have two options. Either the utterance “you ought do *x*” makes an assertion, or it does not. First consider the analysis according to which it is not a genuine assertion. There are a number of different ways to flesh this out. Here, we will consider just one, due to Hinckfuss.<sup>13</sup> To say “you ought do *x*” (as a proponent of a moral fiction) is to assert that you ought do *x*, under the *presupposition* of the fiction and the bridge laws.

---

<sup>13</sup> Hinckfuss (1993)

The situation is no different to the case of fictionalists about mathematics declaring that there is a prime between 2 and 5, despite the fact that they do not hold that there is such a prime number, since there are no prime numbers at all. The fictionalist about mathematics is not committed to there *really* being a prime between 2 and 5 by saying that there is such; the moral fictionalist is not committed to there *really* being something evil about female infibulation by saying that it is evil, any more than we are committed to the existence of hobbits by talking about Bilbo Baggins in the appropriate context. The conversational context in question is the presupposition of the fiction. Using moral terminology, the fictionalist *presupposes* the false theory, and the claims made are appropriate when they follow from the fiction combined with what the fictionalist takes to be true. The function of modifiers such as “*really*” in claims like “there is *really* no good or evil” is to retract the presupposition. So, utterances that look like moral assertions are not genuine *assertions*, but *conditional* assertions. They are assertible when the corresponding conditional (“if the fiction is true then...”) is assertible. On this analysis, in saying “you ought do *x*” you do not assert the conditional, you make a conditional assertion.

This analysis of assertibility is tempting, but it is certainly not the only theory available. There are alternative accounts of fictional assertions in the literature which agree that no genuine assertion is being made. For John Searle, a fictional utterance is a *pretend* assertion, with its illocutionary force suspended.<sup>14</sup> For Greg Currie, it is the performance of another speech act altogether.<sup>15</sup> Such stories still provide the resources to give standards that distinguish appropriate fictional utterances from inappropriate ones.

Kendall Walton offers another account according to which assertoric sentences produced in fictional contexts make assertions.<sup>16</sup> What is asserted on Walton's view, however, is not the same as the sentence's literal content (if any). It is instead an assertion whose content is given by what conditions would need to be met in reality to make the sentence fictionally true—the hearer works out what would need to be in fact true for the statement made to be fictionally appropriate, and in effect takes the speaker to have committed herself to that real-world

---

<sup>14</sup> Searle (1979)

<sup>15</sup> Currie (1990)

<sup>16</sup> Walton (1990), Chapter 10.

condition obtaining. This account of fictional assertion has been taken up by recent fictionalists such as Crimmins (1998) and Yablo (2001).

The final option to consider takes the “fictional utterance” to make a genuine assertion. Here, as elsewhere, we have a range of analyses. According to some, when I say “you ought to  $x$ ”, I am asserting something true, but it is not literally expressed by the surface grammar of the expression. For David Lewis, my assertion has a hidden “it is true in the fiction that” operator.<sup>17</sup> For Edward Zalta, my utterance involves a different mode of predication.<sup>18</sup> Take Lewis’s account as an example. On his account, the *truth* expressed by the claim “you ought to  $x$ ” is the claim “according to the fiction, you ought to  $x$ .”<sup>19</sup> Here, the “you ought to  $x$ ” will be assertible for the fictionalist when, according to the fiction,  $x$  is good. This strategy gives the same *extension* for acceptability as does Hinckfuss’s strategy: “according to the fiction,  $x$  is good” is true if and only if under the supposition of the fiction,  $x$  is good.<sup>20</sup> So, these strategies give the same answers to the assertibility question, while analysing assertions differently.

Finally, if we take a “fictional utterance” to be an assertion with the same content as it would have in literal contexts, we could take the utterance to be straightforwardly expressed by its surface form, and as a result, to be literally false. However, some of these utterances can nonetheless be assertible if we take the assertibility conditions to track, not the content of the assertion itself, but rather, the content of some claim *communicated* by the assertion. When I say “Bilbo Baggins is short”, I am making an assertion that is literally false (*contra* Hinckfuss, who takes me to make a conditional assertion, or Lewis who takes me to make a true assertion of the form “according to *The Lord of the Rings*, Bilbo Baggins is short”).<sup>21</sup> However, it may communicate some other truth (e.g. construction Lewis uses as the content of the assertion) given the context of utterance. If this analysis is preferred,

---

<sup>17</sup> Lewis (1978)

<sup>18</sup> Zalta (1992, 2000)

<sup>19</sup> This differs from the presupposition strategy. To say “ $x$  is good” is to say, under a presupposition, that  $x$  is good. It is not to say *that* under the presupposition of the fiction,  $x$  is good. The scope of the ‘that’ clause is different.

<sup>20</sup> The relevant manoeuvre is *conditional proof*. The conditional “if  $A$  then  $B$ ” is true if and only if under the supposition that  $A$ , “ $B$ ” is true. Of course, different techniques for supposition give conditionals with different behaviours, but this is not our topic here. (See Hinckfuss (1990) for a discussion.)

<sup>21</sup> Recall the distinction between sentence meaning and speaker’s meaning. Here the sentence meaning might be false, but the speaker’s meaning might also be true.

then the distinctive nature of fictional assertions is the assertibility conditions tracking those of the information communicated instead of the literal truth asserted.

We have many options, then, for explaining the appropriateness or otherwise of assertions plucked out of the moral fiction. Just as we have no settled analysis of acceptability for fictional assertion in general, we have no settled account for acceptability conditions in moral fiction. We are confident, however, that Hinckfuss's conditional assertion analysis will explain the difference between acceptable and unacceptable moral utterances, despite the literal falsehood of most moral discourse.

## **2. Stating Moral Fictionalism**

Fictionalism provides a way to deal with the distinctive nature of moral discourse. Moral discourse is useful in many ways, yet theories which attempt to give truth conditions for moral claims founder. Moral claims (at least positive ones—such as the claim that to cause suffering is morally wrong, in general) are, strictly speaking, *false*, just as claims about fictional characters (at least positive ones—such as the claim that Sherlock Holmes lived in Baker Street) are, strictly speaking, false. To state that Sherlock Holmes lived in Baker Street is to state that Holmes existed—but Holmes did not. To state that causing suffering is morally wrong is to ascribe a motivating objective property to a kind of action—and there is no such property. However, in the moral case, these falsehoods are *useful*. For reasons we shall shortly discuss, it is extremely difficult to do away with moral talk.

### ***Defining Moral Fictionalism***

To say this much is to sketch the direction of a moral fictionalism. To define the position we must do more. Here is what we require for a moral theory to be truly described as a *moral fictionalism*.

*The Base Discourse:* Each fictionalism has a base discourse. In the case of moral fictionalism, this discourse must contain no positive moral claims of any kind. This does not mean that the discourse contains *no* moral terminology. After all, we want to say that *literally speaking*, it is not the case that causing suffering is morally wrong. This is a true claim of the base discourse, and it contains moral vocabulary. We must be careful in spelling out the nature of the base discourse.

In saying that the base discourse contains no *positive* moral claims, we mean, at least, that the base discourse contains no claims of the form

$x$  is morally good     $x$  is morally evil     $x$  is morally right  
 $x$  is morally wrong     $x$  is a right     $x$  is a duty

or any of their synonyms. No morally evaluative claim is made by the base discourse. Similarly, the discourse makes no claims of any moral ontology. According to it, there are no *rights*, no *duties*, no *obligations* and no *permissions*. The base discourse must be acceptable to a moral nihilist.<sup>22</sup> It must contain no claims ascribing objective prescriptivity.

*The Fiction and the Bridge Laws:* A moral fictionalism contains a fiction and bridge laws that *do* make positive moral claims, when combined with the base discourse.<sup>23</sup> The fiction may be very rich with many connections between moral properties, or with a rich moral ontology. On the other hand, the fiction may be empty or nearly empty, and the content might be in the bridge laws. The only requirement we make for a moral fictionalism is that morally evaluative claims are given in the combination of the base discourse with the fiction and the bridge laws. So, bridge laws will connect moral discourse to non-moral discourse by way of biconditionals or conditionals of the usual kind, and the fiction will be a moral theory couched in these moral terms.

It follows that the relationship between moral fictionalism and moral discourse differs from the relationship between *modal* fictionalism and *modal* discourse. Modal fictionalism takes possibility and necessity judgements to be literally true, but the traditional possible worlds semantics for modal discourse is only fictionally true.<sup>24</sup> Strictly speaking, modal fictionalism is not a fictionalism about *modality*; it is a fictionalism about *possible worlds*. Our *moral* fictionalism is a fictionalism about morality. We take it that the problematic features of moral discourse affect *all* positive moral claims, and that a fictionalism must treat all of them.<sup>25</sup> The situation is not analogous to modal

---

<sup>22</sup>In fact, this provides a more nuanced technique for specifying the base discourse. It is a discourse acceptable to a moral nihilist. Take Hinckfuss (1987) as a suitable example of an extensively developed moral nihilism.

<sup>23</sup>Well, perhaps. We suppose, for example, that there might be a divine command theory combined with atheism. Then there are no instantiated moral properties, but there would be if there were a God. We leave it to the reader to categorise this view.

<sup>24</sup>Rosen (1990).

<sup>25</sup>A less radical, but arguably still moral fictionalist, position might parallel orthodox modal fictionalism, accepting the literal truth of some positive claims involving right and wrong, good and evil etc (e.g. torturing innocents is evil), and

fictionalism, in which the “incredulous stare” applies to judgements of possible worlds and their denizens. They do not apply to judgements of possibility and necessity.

Before arguing for moral fictionalism, and before defending it against objections, we will illustrate it with a number of examples.

### ***Examples of Moral Fictionalisms***

Moral fictionalism is very general. To give some idea of its generality, we will consider a two different ways it can be developed.

A SIMPLE ACT-UTILITARIANISM: The base discourse is some discourse acceptable to the moral nihilist. The only requirement is it contain the concepts of happiness and actions. For a *bridge law* we have:

The action *x* is *better than* the action *y* if and only if  
the action *x* would cause greater overall happiness than *y* would.

Then the *fiction* can contain a very minimal claim.

An agent *ought to* do *x* only if  
no *better* action incompatible with *x*  
is possible for that agent.

This is a *minimal* moral theory—a very crude act utilitarianism, with two moral predicates, one of relative goodness, and one of obligation. It could be fleshed out more realistically in a number of ways.<sup>26</sup> One distinctive feature of this fictionalism is its minimality. This fiction is logically conservative. The fiction together with the bridge laws may be combined consistently with any base discourse whatsoever. The fiction licences no non-trivial inferences about matters of the base discourse. In this sense it differs from fictionalism about quantum physics, which undoubtedly has macrophysical consequences. There are no such consequences for this fictionalism. The bridge laws and fiction do not license any inferences about agents or actions other than those involving obligation or relative goodness. They may be

---

treating only talk of the ontology of morality fictionally. Bentham (see Ogden 1932) may have held a version of such an ontological moral fictionalism.

<sup>26</sup> See Smart and Williams (1973).

consistently added to any collection of possible non-moral truths. This can be seen as a version of the thesis that no *ought* follows from an *is*.

Not all moral fictionalisms need be logically conservative. Here is a richer moral fictionalism with contingent non-moral consequences.

A SOCIOBIOLOGICAL VIRTUE ETHICS: This fiction grounds the moral in *virtue*. It has an extremely simple, and correspondingly crude bridge law grounding virtue in evolutionary biology.

*x* is a *virtue* if and only if its possession  
contributes positively to its bearer's reproductive fitness.

The moral fiction can contain any plausible theory of the virtues. For example, we could add claims such as those that define virtues more:

*Bravery, humility, honesty, good humour and creativity are virtues.*

This treats *bravery, humility* and the like as moral terms. We could provide more bridge principles connecting these to non-moral terms if you like, such as

*x* is *brave* if and only if *x* is *F*

where the right-hand side of this biconditional is some non-moral description *F* of what is involved in being brave.<sup>27</sup> Once the fiction is so extended, it will almost invariably lose its logical conservativeness. It is a consequence of the fiction and the bridge laws that *F*ness contributes positively to its bearer's reproductive fitness. For most interesting *F*s this is contingent.

Other elements of the fiction might relate virtues and other moral concepts.

*Right action* is that which flows from a *virtue*.

And the fiction can be fleshed out in any number of different ways. We need not detain ourselves with this—the idea has been to give an

---

<sup>27</sup> Of course, not everyone agrees that there are such non-moral descriptions of virtues. No matter. The bridge principles will then be more complex, and perhaps will not fully determine the extension of the predicate “brave” in the fiction.

indication of the division between the base discourse, the bridge principles and the fiction. Neither of these examples are richly developed moral theories. They do not need to be. More extensive theories can be constructed from any developed moral theory. If you have a realist moral theory, it is a straightforward matter to carve up the claims of the theory: Purely moral claims can go in the fictional discourse. Claims relating moral properties to non-moral properties are the bridge laws. Any moral theory may be adopted as a moral fictionalism. This is not to say that constructing an appropriate moral fiction is particularly easy. Questions of moral psychology can be seen as debates over the most appropriate construction of bridge laws.

### **3. Moral Fictionalism and its Anti-Realist Rivals**

#### ***Motivating Moral Fictionalism over Eliminativism***

It is often thought that the natural or inevitable upshot of moral nihilism (the view that all positive moral claims are false, or near enough) is moral eliminativism: the view that we should cease to make the false moral claims. According to this view, if the error theory is true, and much of our ordinary realist moral discourse is irredeemably steeped in error, then consistency demands that we should jettison the contaminated discourse, perhaps in favour of suitably sanitised talk of non-moral preferences and desires.<sup>28</sup> Instead of saying that ‘torturing innocent babies for fun is wrong’ we should say that it is something that we strongly dislike, or strongly prefer not to happen; and perhaps also that we are prepared to stop it and punish those who do it. Likewise, rather than saying that ‘we have a duty to help those in need’, we should say that we strongly prefer that we help those in need, and perhaps also that we’re prepared to encourage others to do the same.

Some moral nihilists, perhaps most notably Ian Hinckfuss, have welcomed both the truth of nihilism and the eliminativist response to it.<sup>29</sup> There lies liberation from the shackles of a pernicious and

---

<sup>28</sup>Simon Blackburn writes “Mackie did not draw quite the consequences one might have expected from this position [the error theory]. If a vocabulary embodies an error, then it would be better if it were replaced by one which avoids the error. We could better describe this by saying that our old, infected moral concepts or ways of thought should be replaced by ones that serve our legitimate needs but avoid the mistake.” (Blackburn (1993a), p. 149.)

<sup>29</sup>Hinckfuss (1987), esp. Part 3. Hinckfuss claims that many acts of war, terror and genocide would most likely not have occurred had the perpetrators not believed that right was on their side.

oppressive ideology, which has been the cause of little other than much avoidable human misery. But it is far from obvious that eliminativism is a price that we should be happy to pay. Fictionalism offers itself as an alternative: a way of retaining the realist talk, without the undesirable commitments. But if fictionalism is to be an alternative, the fictionalist must answer the eliminativist challenge: what reason do we have for continuing to use the language of realist morality, given that it is founded in error? The fictionalist's answer, in brief, is that realist moral discourse should be retained, even though it is strictly speaking false, because it is useful. (The 'should' here is a pragmatic should, not a moral one). It is useful to engage in the pretence that positive moral claims are true, even while believing that they are literally false. But how might the employment of moral talk be useful, if such talk does not state the truths it purports to state?

One obvious advantage of fictionalism over eliminativism is psychological convenience. We've grown deeply accustomed to saying that things are right and wrong and to talk of duties, virtues, rights, justice and obligations. These concepts pervade ordinary thinking and discourse. Perhaps it is not psychologically impossible to give up these ways of talking and thinking in favour of talk only of non-moral preferences and actions, but it would be difficult and inconvenient. Certainly, giving up moral discourse is not like giving up a relatively isolated and rarefied concept in a scientific theory, such as the concept of phlogiston. Eliminativism about moral discourse would force great and wide ranging changes to our patterns of speech and thought on much the same scale as would eliminativism about folk psychological concepts of the sort famously proposed by the Churchlands.<sup>30</sup> Ceasing to talk of goodness and badness, rightness and wrongness, duties, justice, and obligations would be much like ceasing to talk of people having beliefs, desires and emotions: possible, perhaps, but not an easy thing to do; and certainly not a consequence to be embraced lightly. We think that this is no small motivation for seeking a way of retaining the talk, without committing the error.

A second advantage of fictionalism is that it enables us to avoid raising complicated meta-ethical issues every time we wish to discuss an applied ethical question. Suppose, for example, a reporter asks for our opinion about the justice of the new inheritance tax; or suppose a friend

---

<sup>30</sup> See, for example, Churchland (1988).

is agonizing over whether she should have an abortion and asks our advice about what she should do. When “speaking with the vulgar”, fictionalism enables us to avoid being sidetracked into a lengthy meta-ethical discussion about the status of the terms of discussion and the truth of the error theory. We do not have to reply, rather unhelpfully, to our friend, “well, I don’t really believe that you should do anything since I don’t really believe that there is anything that ever should or should not be done; nor do I believe that it would be wrong for you to have an abortion, or right for you to have one for that matter, but...” If we can talk as if we have a positive opinion on these matters, we can get to the point of the discussion right away.

A third advantage fictionalism may have over eliminativism is that of expressive power. We have already hinted at this in the case of mathematical deductions. A proof, taking a detour from the empirical, through the numerical, and back into the empirical, may be much more succinct than a demonstration proceeding directly. The same is the case for *assertions* that make use of the fictional discourse. To say that there are no prime numbers between 23 and 29 is much more succinct than any equivalent nonmathematical expression. To say *that* without using the fiction we would have to say something like this: if you have more than 23 but fewer than 29 things, then there is always a way to divide these things into a number of classes of equally many things. This is more cumbersome, and it still uses a mathematical expression, that of a *class* of things. You could use a geometrical paraphrase: if you have more than 23 but fewer than 29 things then you can arrange these things in a rectangle with the same number of objects in each row, where this rectangle either has more than one row and fewer rows than there are objects. Again, this is more complex, and it still uses possibly unacceptable elements of discourse, like “rectangle,” and “row.”

Examples from the metaphysics of properties are also familiar. A nominalist with a fictionalist disposition can *use* certain false claims about, such as “red is more similar to orange than it is to green” or “there are two important properties shared by Napoleon and Alexander the Great” to succinctly communicate truths. A great deal of the debate over nominalism concerns whether such claims can be translated into nominalistically acceptable language.<sup>31</sup> Perhaps they can, or perhaps

---

<sup>31</sup> Pap (1959), Jackson (1977).

they cannot. In either case, the fictionalist can agree that they commit themselves to something true in virtue of expressing the falsehoods. Perhaps these truths can be expressed in the base discourse, but perhaps they cannot.

Similar considerations may apply in the moral case as well. There are sentences with moral vocabulary, which we use to imply things about non-moral features of the world, but it seems difficult to say exactly what those features are in non-moral terms. Take a common sort of moral claim such as ‘the property rights of the farmer outweigh the rights of the environment’. Exactly what non-moral features of the situation does talk of ‘rights’ capture? Suppose that we want to say that the duty to stop and attend to the victims of a car accident is more important than our duty to keep our commitment to meet our friend for lunch. Again, what non-moral features of the world are implied by talk of ‘duty’ here? The fictionalist can avoid the difficulties of finding an appropriate literal non-moral paraphrase. We can say that it’s *literally* true that if there are rights then the property rights of the farmer are more important than the rights of the environment. This literal truth has empirical consequences, notwithstanding any “is-ought” gap. It is a non-normative consequence of this truth that there *are* farmers—taking a trivial example—and this is a contingent truth that requires no moral vocabulary to state.

A fourth advantage of moral fictionalism over eliminativism has to do with its capacity to salvage the important role moral discourse is widely thought to play in co-ordinating attitudes and regulating interpersonal conflict in cases where people disagree about what they are to do, especially where collective action is needed or the proposed actions of different people interfere with each other. The institution of realist morality has taught us to discuss the resolution of such matters in terms of rights, duties and obligations, and to abandon these frameworks for discussion would deprive us of many of the procedures and tacit understandings that provide a well-established framework within such discussions take place. Realist morality tells us that there are correct answers to questions about what the morally appropriate decisions and outcomes are, and that when there is genuine disagreement about what ought to be done at most one of the parties can be correct. A realist commitment to the institution of discussing rights and duties can thus be made sense of, provided it matters to parties to the dispute that they do the right thing (or even are seen to

be doing the right thing). Naturally, moral debate does not always succeed in achieving an agreement in attitudes, especially in cases where there is little agreement about fundamental moral principles. These patterns or conventions of talking and thinking when deciding on collective action or resolving practical conflicts are ubiquitous and important in maintaining any social relationships: and it seems *prima facie* an advantage to be able to hold onto these.

The problem for the fictionalist is to explain how these practices of dispute could retain their force once realism is abandoned. It is easy enough to see why a fictionalist might retain portions of this practice when dealing with realists, since arguments about obligations or justice will have *ad hominem* force. But suppose that all the parties to the debate are fictionalists. Why might people change their preferences about how to act on the grounds of what is true in a story? There is a two-pronged answer that the moral fictionalist can offer, corresponding to the two main alternatives moral realists have in explaining why someone's preferences might be changed as a result of their changing their moral beliefs. One reply would connect non-moral preferences and what is true in the fiction via internalist bridge laws (though care must be taken in stating these). If the fiction is set up in such a way that it is guaranteed that something's being good-in-the-story that the people engaged in the story have certain non-cognitive attitudes towards it, then coming to realise that some course of action does have certain moral properties according to the story should prompt the realisation that the action is one that the agent has certain attitudes towards.

Or the fictionalist could tell an externalist story: it might be the case that, by and large, people contingently want to bring about situations which are true according to the fiction. According to the externalist, no mere cognitive belief alone will affect people's preferences, but that does not mean that people may not alter their preferences to reflect what is true in the fiction. For people may care that their attitudes are in line with what is true in the fiction, and so alter their non-moral preferences accordingly. Further explanation of why one might care that one's non-moral preferences reflect what is true in the fiction will depend on the details of the particular fiction that is being employed. (Presumably the most common will involve the fiction designating as good or right events or actions which are in fact widely desired – we expect that relying on some intrinsic appeal to pretending to be moral will not be sufficient once the stakes become serious). We will say more

about the nature of moral disagreement among the self-consciously fictionalist in section 3.

These are not the only reasons that might be advanced in favour of fictionalism over eliminativism. Richard Joyce points to the benefit of continuing to pretend that we have genuine moral commitments in overcoming weakness of will.<sup>32</sup> An agent engaging in the pretence that a certain course of action is the right one—an action which ‘simply must be done’—is more likely to withstand the temptation of doing otherwise, than an agent who simply judges it to be in her best interests to perform it. One might also think that moral vocabulary is required in childrearing.<sup>33</sup> Concepts such as right and wrong, good and bad, virtue and vice might play an indispensable role encouraging children to behave in certain ways, and to refrain from behaving in others.

We are happy to allow that it is an open question whether these reasons are good enough for keeping moral discourse. It may be that these reasons obtain in some situations but not in others; in which case, a tactical fictionalism is an option: we keep the discourse until a better set of categories is constructed, or until the youth are reared, or until years of habits have curbed the wandering will. We do, however, regard these considerations as providing good enough motivation for exploring a fictionalist alternative to eliminativism. At the least, fictionalism is a preferable alternative to the sort of ‘inconsistent nihilism’ sometimes explicitly embraced by error theorists who find themselves unwilling or unable to refrain from making the positive moral judgements they take to be false.<sup>34</sup> Inconsistent nihilists are committed inconsistently. Their theory dictates that moral claims are false. Some impulse urges them to take these claims to be true. A fictionalist account takes these urges and allows them to be expressed. They are not literally true, but we may act as if they are. We are not committed to inconsistency, so if commitment to the fiction assuages the desire to take moral discourse to be true, fictionalism is superior to inconsistent nihilism.<sup>35</sup>

---

<sup>32</sup> Joyce (2001), ch. 7.

<sup>33</sup> Charles Pigden has put this point to us in discussion.

<sup>34</sup> Pigden (1991) describes himself as a reluctantly ‘inconsistent nihilist’.

<sup>35</sup> The “protected addition” strategy, discussed earlier, can guarantee that fictionalist discourse is consistent, given that both the fiction with bridge laws and the base discourse are both consistent. There is no *extra* problem for consistency for the fictionalist.

Fictionalist approaches to an area of discourse share a virtue alien to related approaches, like quasi-realism. The combined theory of base discourse, bridge principles and fiction has a uniform interpretation. As a result, the combined theory is closed under logical consequence. There is no difficulty in explaining the validity of reasoning which detours through the fiction. There is no special status given to the fiction, no different kind of ‘truth’. There are no Frege–Geach problems for this view.<sup>36</sup> The closure of the combined theory under logical consequence means that fictionalism has a straightforward explanation for the appropriateness of reasoning crossing over distinct categories such as mathematics, the unobservable or the moral. The reasoning is valid since the premises and the conclusion are all treated realistically. Of course, some of the premises are false, but this is no impediment to validity. Valid reasoning can use false premises or deliver false conclusions. It cannot, however, appeal to premises that are not *truth-apt*. Fictionalism is ideal for discourses that are amenable to logical reasoning, such as morality. It is to expounding these and other advantages in more detail that we now turn.

### ***Advantages for moral fictionalism over other non-realist approaches***

We will not defend moral fictionalism directly. Rather, we will compare moral fictionalism with other well-known non-realist approaches. These include traditional non-cognitivist theories (such as those of Ayer, Stevenson and Hare<sup>37</sup>), however we will focus on quasi-realism—the quasi-realism of Simon Blackburn—because the quasi-realist approach is closest to the moral fictionalist’s in seeking to keep as much realism as possible, whilst avoiding realism’s problems. We think that everything desirable that quasi-realism can do, moral fictionalism can do as well or better; and moreover, moral fictionalism does so in a way that avoids the most pressing problems with quasi-realism.<sup>38</sup>

---

<sup>36</sup> See Scheuler (1988) for a succinct statement of this range of problems for anti-realisms.

<sup>37</sup> Ayer (1952); Hare (1952, 1995); Stevenson (1944, 1963).

<sup>38</sup> In a conference presentation at the Australasian Association for Philosophy meeting in Hobart 2001, David Lewis argued that quasi-realism was best interpreted as a variety of fictionalism, rather than a distinct position. We think that quasi-realism might be improved by becoming fictionalism, but as our remarks in the text indicate, we do not think it counts as fictionalism as it stands.

One of the primary motivations for quasi-realism is that it provides a reconciliation of the apparent objectivity of moral judgements with internalism: the idea that, necessarily, moral beliefs or knowledge or facts give us reasons for acting morally, either by giving us motivations or at least by making demands on us insofar as we are rational. The appeal of internalism lies in the intuition that moral beliefs are action-guiding and that someone who has a genuine moral conviction cannot remain indifferent to that moral concern, at least absent some mental or rational defect. The problem for realist accounts, as we sketched in our introduction, is that they have trouble telling us how there is some kind of objective fact in the world which produces motivation in us when we come to hold an opinion about it or, at any rate, which motivates us insofar as we are rational.

Quasi-realism accounts for the connection between our talk that seems to presuppose the objectivity of moral facts and reasons for acting by arguing that we build up our apparently objective moral talk from the basis of our sentiments. Quasi-realism is a descendent of expressivism: moral claims are not reports of our sentiments, rather they prescribe what is appropriate to say morally. How? One way is via minimalism about truth: once we have assertion conditions with a fairly stable set of conditions about what is appropriate to assert and what is not appropriate to assert then we have truth conditions. The most perspicuous story about what is assertible is to be given, not in terms of moral properties attaching to events, outcomes, people etc in the world, but rather in terms of our attitudes and how they might change in various circumstances. For example, let us suppose that “abortion is wrong” is assertible just in case people with mature, healthy sensibilities who have been given all the factual information would disapprove of abortion. If we are minimalists about truth, the fact that this assertion condition is reasonably well-behaved (however that is spelled out) means that such assertions are either minimally true or minimally false. Let us suppose that abortion is something that people with mature, healthy sensibilities who have been given all the factual information would disapprove of, then since “abortion is wrong” is minimally true (and so true in some sense), abortion is wrong.

Of course, quasi-realists do not have to rely on minimalism about truth. Early versions of quasi-realism instead gave assertion conditions for sentences applying the truth predicate to moral sentences. These assertion conditions will also be in terms of facts about what people’s

attitudes are or would be. Once this projection is completed, we can agree that abortion is indeed wrong and we can even agree to claims like “the property of wrongness attaches to acts of abortions” (again, assertion conditions for this will be quasi-realist ones). Further quasi-realist constructions allow for moral inferences, for claims about what would be moral in other circumstances, and so on.

It is clear how the quasi-realist provides for a necessary connection between the obtaining of moral properties and the possession of certain attitudes. After all, in the abortion case above, when abortion is wrong, people with mature healthy sensibilities who have been given all the factual information disapprove of it. (It can be more difficult to state the connection once quasi-realism is extended to modal claims involving morality, since it might turn out that abortion would still be wrong, even if people with mature and healthy sensibilities had approved of it. Blackburn in *Spreading The Word* is careful to construct the relevant counterfactuals so that if our sentiments had been different still much of the moral truth would not be.<sup>39</sup>)

Moral fictionalism can also easily accommodate internalism if desired. We can guarantee that it will be true that, according to the fiction, some state of affairs concerning morality obtains just in case people have certain motivations or pragmatic reasons by including bridge principles that, in certain circumstances, take one from facts about motives or reasons, on the one hand, to facts about moral truth or beliefs genuinely about morality, on the other, and vice versa. Which bridge principles are employed will depend on how exactly the fictionalist cashes out their internalist commitments—we do not propose to determine what the most plausible internalist position is here.

Another advantage which quasi-realists claim over realists concerns the epistemology of moral truth. Since, for the realist, moral beliefs represent a domain of moral facts in the world, the realist faces the challenge of explaining how it is that we come to know or be acquainted with these facts. For the quasi-realist, on the other hand, the explanation of how we come to know about the moral facts is straightforward. Since moral facts have their basis in our sentiments as they actually are, or as they would be under suitably idealised circumstances, we come to know about these facts simply by coming to

---

<sup>39</sup> Blackburn (1984), p 219.

know about ourselves. The moral fictionalist has an equally straightforward way of accounting for our knowing when to make moral assertions. All we need to know is what is in the fiction; and there is nothing mysterious involved in being able to know what story we are telling.

Yet another advantage the quasi-realist claims over the realist is that of being able to explain the supervenience of the moral on the non-moral. A challenge the realist faces, if there are these independent moral facts, is to explain how these moral facts are related to the non-moral ones: in particular, to explain the alleged conceptual truth that there can be no possible difference in the moral without a difference in the non-moral. Two situations that are exactly alike in all non-moral respects must be alike in moral respects. The explanation the quasi-realist offers is that the moral supervenes on the non-moral because the appropriate sensibilities evaluate situations that are alike in all non-moral respects as alike in all moral respects. Guaranteeing the appearance of supervenience is just as easy for the fictionalist. The fictionalist who wants to ensure no difference in the assertibility of moral claims without a difference in the non-moral facts just needs to specify the bridge laws in such a way that they do not permit any such difference.

As well as advantages over realists, quasi-realists claim to have advantages over their expressivist ancestors in dealing with a cluster of difficulties for anti-realism which have come to be known as the Frege-Geach problems<sup>40</sup>. The problem is that if moral talk did not literally make claims, but instead did something else (merely expressed feelings, or issued disguised commands, for example), it would not make sense to embed them in complex statements with truth-functional connectives (and, or, not), or to embed them in conditional sentences, or employ them in deductive or inductive inferences. If “taxpaying is good” is just approval of taxpaying, “it is not the case that taxpaying is good” should make no more sense than “it is not the case that whoopee for taxpaying!” Should we treat “you must not lie” as a command not to lie, then “if you must not lie, you must not cheat either” would be as nonsensical as “If do not lie then do not cheat either!” is. And the inference “if you must not lie, you must not cheat either”, “you must not lie”, therefore “you must not cheat” would also be virtually unintelligible, whereas it is clearly in order, if not quotidian.

---

<sup>40</sup> For the classic statement of the problem, see Geach (1965).

Different quasi-realist approaches vary about the details of the proposed solutions, but in all cases they provide ways to handle these complex constructions by producing assertibility conditions for them from a basis in our attitudes. We shall not discuss the details of these different solutions here, suffice it to note that all of these strategies are extremely complicated, and many are sceptical that any of them ultimately succeed.<sup>41</sup> Moral fictionalism, on the other hand, despite being a non-realist strategy, finds these challenges as simple as the realist does. The semantics offered for moral language, after all, is the same semantics as that which the realist offers.

The quasi-realist's ability to assign truth and falsity to moral claims (even if they are minimal truth and falsity) also allows them another advantage over traditional non-cognitivists: they are able to handle moral disagreements more easily. When we disagree, we typically take ourselves to be right, our opponents to be wrong, and one of us to at least be asserting the negation of something the other has asserted. However, if moral claims are mere expressions of attitude, or prescriptions, they cannot be used to disagree in this way. If you express loathing for ice cream and I express sincere appreciation of it, we are not necessarily disagreeing in the normal sense, even if we are in some sense not in harmony: for I need not think that you are making any mistake, or believe something I reject—after all, tastes may sometimes just differ. If I tell Igor to shut the door, and you tell him to leave it open, we are in conflict in a sense, but again neither of us need to think the other has made a mistake, or that either of us has made assertions at all. Quasi-realists, equipped with their (quasi) truth conditions, on the other hand, can take disagreeing parties to be making assertions, and one to assert the negation of what another asserts. After all, if one asserts something which is true just in case what the other asserts is false, they have a classic case of disagreement on their hands. Of course, sometimes when there is a genuine disagreement both sides are wrong (consider the disagreement between a creationist and a Lamarckian about evolution), but a genuine disagreement in which neither side is wrong is more puzzling.

---

<sup>41</sup> To get a sense of the complexity involved in such quasi-realist treatments see, for example, Blackburn (1993b), pp. 193-97. For an argument that Blackburn's proposed quasi-realist solution cannot succeed see, for example, Schueler (1988). Blackburn's "Attitudes and Content" (1993b) is in part an attempt to reply to Schueler's objections.

The fictionalist can do just as well in dealing with genuine moral disputes. Take, for example, the dispute between the fictionalist and the realist. The fictionalist takes them to have a genuine, substantive disagreement about whether anything is right, or good, or morally forbidden etc, and the fictionalist has a literal mode of talking for expressing the disagreement. Since the fictionalist need not be revisionary about the meaning of our moral talk, the realist too can agree on what the dispute is about (though not about which one of them is in fact correct, of course). Fictionalists can also characterise disagreements between rival moralisers in the same way the realist does: while the fictionalist will normally think both parties are mistaken, to some extent, she does not face any problem in allowing they are literally making assertions, nor that those assertions function in the way it is standardly thought incompatible claims do.

It may be a slightly trickier matter to say what it is that two fictionalists asserting apparently incompatible moral sentences are disagreeing about. They will at least have a disagreement about the content of the relevant fiction: but whether this is the subject matter of their disagreement, or whether the disagreement itself is better identified as being elsewhere (e.g. in what they take to be true which they are communicating by literally uttering the falsehoods, if this is distinct from a claim about the fiction) will be a matter which different fictionalist theories might well answer differently. Normally, however, we suspect identifying a matter of genuine dispute when there is apparent moral dispute will be possible.

The non-cognitivist and the quasi-realist face yet another Frege-Geach problem for propositional attitudes. We ordinarily talk as if people have propositional attitudes towards moral claims. We say things such as “I believe that abortion is right” and “Mary wants abortion to be morally permissible”. As well as beliefs and desires, there are hopes and wishes and suspicions and fears. The whole point about propositional attitudes is that they are attitudes to propositions. A belief is correct just when the proposition believed is true; and likewise a desire is satisfied when the content of the desire is satisfied. But, on non-cognitivist stories, there is no proposition that moral claims express. “I believe that Hooray for abortion”, for example, doesn’t make sense. Our propositional attitude talk seems realist: there seem to be facts in the world which our moral beliefs represent (or misrepresent, as the case may be); and our

moral desires seem to be aimed at bringing about certain states of affairs in the world.

The quasi-realist strategy is to preserve our propositional attitude talk, using methods similar to the ones employed in the case of connectives, conditionals, inference, etc. . The quasi-realist has at least two general options here: one is to provide assertibility conditions for propositional attitude ascriptions in terms of the sensibilities of the person in question, or what those sensibilities would be in certain conditions; the other is to provide the truth conditions for the propositions ascribed through one of the previously mentioned methods (e.g. moving from assertion conditions to truth conditions by way of minimalism about truth), and then to follow the realist in saying that *X* believes some moral proposition *p* just in case *X* takes the relevant truth conditions to be satisfied, *X* desires that *p* just in case *X* seeks the satisfaction of the relevant truth conditions, and so on.

Both routes are notoriously problematic. It is very controversial whether it is possible to deliver “minimal” propositional attitudes from minimal truth conditions in this latter way<sup>42</sup>; and the former route faces the same worry about complexity mentioned above. The complexity may become even worse when the quasi-realist comes to try to provide the assertion conditions for nested propositional attitude ascriptions: for example, “Tom believes that Mary believes that abortion is permissible”. Is it that Tom believes that the assertion conditions for Mary to assert “abortion is permissible” (or some translation) are met? But surely that is rather what it is for Tom to believe that “abortion is permissible” is assertible for Mary. Is it that Tom believes that Mary believes that the assertion conditions for “abortion is permissible” are met? Presumably not – perhaps Mary could believe abortion was permissible even if she did not think that the assertion conditions for that claim were met (perhaps she has unusual beliefs about assertion, or perhaps she does not realise that she does so believe in the first place). Is it that Tom is such that if he were to utter “abortion is permissible” in Mary’s circumstances the assertion conditions would be met? No, for Tom might be that way and have no idea that “abortion is permissible” would be assertible for him were he in Mary’s position, through ignorance of Mary’s position, for example.

---

<sup>42</sup> For an example of considerations suggesting a non-minimal account of belief contexts needs to be offered, see Jackson, Oppy and Smith (1994). For one response, see O’Leary-Hawthorne and Price (1996).

Perhaps the quasi-realist following the route of providing assertion conditions for propositional attitude ascriptions has some superior conditions to offer: if so, we await them with interest.<sup>43</sup>

In contrast, for the fictionalist, propositional attitude ascription is as straightforward as it is for the realist—at least, insofar as people are realists about the propositional attitudes. Here it is useful to distinguish between descriptive and prescriptive varieties of fictionalism. As a descriptive thesis, fictionalism claims to be an accurate description of our current moral practices; as a prescriptive thesis, fictionalism claims that, although we are not currently fictionalists about morality, we pragmatically ought to become fictionalists. Both of these are possible fictionalist positions—as are combinations of these. So, insofar as fictionalism is a prescriptive thesis (because people are currently realists in their moral thinking), the fictionalist has an easy time of it explaining what is going on when we when make propositional attitude ascriptions involving ethical propositions. When people claim, for example, that “Mary believes that abortion is wrong” they are claiming just what they seem to be claiming: that Mary represents the world so that abortion has a certain moral feature, namely, being wrong. Similarly, when one desires that sleeping in is morally obligatory one desires that there be a fact—that the Moral Law enjoins that one sleep in. There are certain complications when we come to ascribing propositional attitudes about moral matters when the people involved are fictionalists, since even though e.g. a pro-life fictionalist will not literally believe abortion is morally wrong, still there is a motive to fictionally assert of her that she believes abortion is wrong and this fictional assertion is not about morality per se, but about someone’s beliefs, which is a subject matter the moral fictionalist may not in general wish to be a fictionalist about.<sup>44</sup>

---

<sup>43</sup>Things are so much worse for the quasi-realist who attempts to be a projectivist about propositional attitude ascriptions (such as Blackburn). We know of no quasi-realist treatment of nested propositional attitude ascriptions. Blackburn himself simply points to the availability of what he terms a “fast-track” strategy in this sort of context. (Blackburn (1993b)). The core idea is that the quasi-realist need not answer the Frege-Geach sort of problems by ‘honest toil’ i.e. by providing us with the details of a quasi-realist account for every propositional context that seems *prima facie* problematic for the quasi-realist. Instead, the quasi-realist can simply make sufficient general remarks to motivate the quasi-realist view of ethical discourse; and, given the plausibility of this general explanation, we can then take it on faith that the quasi-realist can come up with adequate answers to the more specific problems for quasi-realism. We think that the fast track strategy is unlikely to do much to assuage the doubts of those who are not already committed quasi-realists.

<sup>44</sup> See [footnote deleted for anonymous review].

A very familiar objection to traditional non-cognitivist approaches (such as those of Ayer, Hare and Stevenson) is that they cannot account for the apparently realist and cognitivist elements of our moral practice: the thought that we can get the answers to moral questions wrong, our concern to get the answers right, and the intuition that the truth or falsity of a moral claim is not simply given by what we are inclined to believe or desire. Moral fictionalism holds a substantial advantage over these more traditional non-cognitivist approaches by remaining cognitivist and thus closer to the way we think; and it maintains this advantage over its more contemporary quasi-realist rival to the extent that it is plausible to construe our ordinary moral notions as involving full-blooded belief and not merely quasi-belief. Moral fictionalism earns its cognitivist credentials in two ways: first, it retains the idea that moral claims are truth-apt (it is just that all positive ones are false); and second, it takes moral claims to express genuine, full-blooded beliefs.

As we mentioned in our introduction, it is a commonplace that the moral truths are objective: the rightness or wrongness of bull-fighting, for example, is not a matter of what we are inclined think or feel or want. Anti-realist approaches famously violate this commonplace. Traditional non-cognitivist approaches violate it by denying that there is any truth to be had at all on moral matters: there are no moral facts for us to get right or wrong and moral claims are subjective expressions of our attitudes or (equally subjective) disguised imperatives. Quasi-realist approaches may at least admit that there is truth to be had, but the moral truths may turn out to be no more objective—at least, when we are considering non-counterfactual moral claims e.g. abortion is wrong. These moral claims are made true (or false) by facts about people's attitudes, or perhaps by facts about what people's attitudes would be under suitably idealized conditions. Subjective truth conditions may be plausible in some domains (the truth-makers for sentences about our attitudes, for example, may be facts about our attitudes), but ordinary thinking tells us that this is *not* the case with moral sentences. If abortion is wrong, it is not wrong because actual or suitably idealized people disapprove of it; rather, insofar as appropriately sensitive, informed people disapprove of abortion, it is because abortion is wrong. Perhaps this commonplace can be given up. But giving it up would be a significant cost to our ordinary notions. Fictionalism, on the other hand, does not face this cost. It does not deny

that moral concepts are concepts of something objective and mind-independent.

Moral fictionalism also seems a more stable position than its quasi-realist rival, which veers a narrow and slippery course between subjectivism, expressivism and non-reductive realism. Quasi-realism is at least initially presented as a species of expressivism: moral claims are supposed to be subjective expressions of attitude. But when it comes to evaluating counterfactuals of the general form ‘if I had felt differently, then such-and-such would have been assertible’, it turns out that the acceptability or assertibility of moral claims does not depend at all on the possession or not of the relevant attitude. As Blackburn himself says “...it comes out *false* that if we had thought or felt otherwise, it would have been permissible to kick dogs”<sup>45</sup>; “... opinion is irrelevant to the wrongness of kicking dogs”.<sup>46</sup> Given that a truth-apt sentence is presumably only appropriately assertible if it is true, we seem to have the conclusion that “kicking dogs is wrong” is both a sentence which is supposed to be an expression of attitude, and its appropriateness has nothing to do with the attitudes of the person who utters it. If this is not just outright inconsistent with the initial expressivist claim that moral claims are expressions of attitude or opinion, then it is at the very least a deeply uneasy bedfellow. Fictionalists, on the other hand, can give an up-front account of what the truth of the matter is: moral claims express beliefs which are truth-apt, but the positive ones of which are false. Fictionalists can be straightforward about why we can make these claims even though we believe them to be false: they are useful. We are all very familiar with manipulating fictions, and it is comparatively unmythical what that involves.

#### **4. Challenges for Moral Fictionalism**

Moral fictionalism faces challenges. Some of these can be answered, some are invitations to further development, and some are objections that may reveal aspects of it which are unpalatable and even fatal.

The first we have already touched on: is moral fictionalism to be construed as a descriptive claim, that in fact what people are doing when they are moralising is engaging in a pretence of morality, using a system they take to be false but useful? Or is it prescriptive: for one

---

<sup>45</sup> Blackburn (1984), p.219, fn. 21.

<sup>46</sup> Blackburn (1984), p.219.

reason or another we *should* employ a moral fiction and continue to assert apparently morally committed claims while withholding belief from them? Or is it a mixture? (One could hold both claims if one thought current practice was indeterminate or variable, or if one thought that what is to be recommended is not entirely clear-cut).

Different fictionalists can go in different directions on this point. We feel, however, that the descriptive option is unlikely to be plausible, unless it is in the context of a global argument that we are fictionalist about many areas in which we are apparently pre-theoretically realists. Hinckfuss's story of presuppositions which are not believed provides one story that might be told by those prepared to suggest that we have been fictionalising all along without being aware of it.<sup>47</sup> We believe that moral fictionalism is more attractive as a prescriptive doctrine: not that we do, but that we *should*, employ moral fiction. This "should" is not a *moral* should; rather, it is some sort of pragmatic "should": given the interest we have in co-ordinating our actions, inculcating and supporting our deeply-held preferences, and perhaps other desires we have, we will be well served to morally fictionalise. Or so says the moral fictionalist, at least. We have already discussed why a moral fictionalist might say this.

Another worry that we have previously mentioned is whether a moral fiction with its associated bridge laws will be a conservative extension of our non-moral talk. The connection between our apparently moral commitments and our commitments in psychology, politics or our theory of normativity are many and varied. It might seem risky to accept the non-moral consequences of these theories put together with the large and complicated falsehood that is the moral fiction.

Moral theories are, as a rule, not comprehensive enough or formulated in a rigorous enough way for results like Field's conservative extension results to be derived by a moral fictionalist. Some fictionalists may wish to have the content of even the core of the fiction depend on contingent matters of fact, and those fictionalists may not be interested in a fiction which is a logically conservative extension of non-moral theories. Mounting a convincing case that the theory is a truth-value preserving conservative extension is difficult, if it is rich enough to rival full-

---

<sup>47</sup> Hinckfuss (1993)

blooded theories of morality, given the incomplete development of much of our moral attitudes.

One way for a moral fictionalist to respond to this challenge is to present a comprehensive theory for inspection, and demonstrate that the theory is a conservative extension of our non-moral theories, or to show why we might have good reason to expect so. Less ambitious responses are also available. One is to find partners in guilt: it is open to the moral fictionalist to claim that their claim that their moral fiction, considered together with non-moral truths, will not permit any non-moral falsehoods to be derived, is on as sturdy a ground as the moral realist's claim that their moral theory, together with the moral truths, will yield no non-moral falsehoods as consequences. After all, given a moral theory that specifies the connection between the moral and the non-moral in a thorough way, the fictionalist can take that theory to instead be a fiction, merely by altering their attitude to the core theory and the specifications of its connections with non-moral theories. In a case where the fictionalist does so, the fiction will fail to be a truth-value preserving conservative extension of the non-moral truth only when a falsehood can be derived from the moral theory together with the non-moral truth: a condition which would make the moral theory untenable for the moral realist. So if we have reason to accept that the moral realist's moral theory is a truth-value conserving extension of the non-moral truths, then moral fictionalists can avail themselves of that same reason to accept it as a truth-value conserving extension.

The realist need not accept this *tu quoque*, as the discussion following van Fraassen's defence of this strategy shows. This move will be most tempting to those moral realists who adopt an epistemology which is more than merely coherentist: for if the only sort of evidence a moral realist has for his moral claims is that they are consistent, and consistent with his non-moral beliefs and evidence, then he will find it difficult to deny that the fictionalist may use this evidence as *some* evidence that the moral belief, with the non-moral truth, does not imply non-moral falsehood.

However, the moral fictionalist is not without a counter to even the moral realist who takes himself to have some more-than-coherentist evidence for a positive moral claim. The fictionalist can again attempt to parallel the realist, by claiming that what the realist takes to be evidence for the claim's truth, the fictionalist takes to be evidence for

the claim's status as not implying a non-moral falsehood when conjoined with the non-moral truth. This is analogous to van Fraassen's claim that the fictionalist can employ the evidence a realist takes to indicate the existence of unobservables of various sorts to instead be evidence for the *acceptability* of the relevant claims about unobservables. Van Fraassen argues that the evidence which a realist takes to be grounds for an "inference to the best explanation" can instead be taken to be rational grounds, not for belief in the truth of a theory, but rather for the belief that the theory is empirically adequate: that the theory has no false consequences about the observable realm.<sup>48</sup> Since a realist must take those claims he holds to be true to be at least acceptable as well, the realist cannot object to a fictionalist making the inference to acceptability, even if she declines the further inference to the truth of those claims.<sup>49</sup>

Another source of concern about moral fictionalism, as with fictionalism in many theoretical areas, is of a Quinean, pragmatist or holist nature.<sup>50</sup> The fictionalist proposes dividing our useful utterances which apparently tell us about the world into two categories: one to be taken at face value as really telling us how things are, and another which is mere pretence. Why should there be this discontinuity? Furthermore, since the moral fictionalist demands that a distinction be drawn between the two kinds of talk, where is the line to be drawn?

Neither of these challenges lends itself to an obvious and straightforward response on the part of the moral fictionalist, we believe. Nonetheless, there are things the moral fictionalist can say to both questions. One thing that can be said in response to the first question is that we have different kinds of justification for the different theories that we use. The most common is in terms of evidential considerations, e.g. we think that the theories provide the best explanation, or that the theories are best supported by the observations available. On the other hand, there are theories or bodies of information that we manipulate for different purposes, e.g. we tell the story about Sherlock Holmes to entertain, not to convey information about the behaviour of London detectives. We could concede to the holist that theories employed for the first sort of reasons stand or fall as a whole,

---

<sup>48</sup> See van Fraassen (1980), pp. 19–21

<sup>49</sup> Van Fraassen makes this point about the position of the fictionalist most clearly in his discussion of Vaihinger's position in van Fraassen (1980), p.36.

<sup>50</sup> Putnam (1972)

and have important interconnections and that it is difficult to draw principled evidential boundaries between theories of that sort. But that is not to say that we cannot draw distinctions between theories of the first sort (intended to be literally true) and theories employed for other reasons, e.g for psychological convenience, or simplicity of exposition. Physicists talk about ideal gases and frictionless surfaces, but not even the Quinean thinks we have to literally believe in them.

In response to the second challenge, the moral fictionalist might point out that it is widely thought that it is possible to demarcate the moral from the non-moral terms. It may be that the boundary is vague—in which case there will be vagueness about which claims to treat literally and which claims to treat fictionally. But in any case, *everyone* faces the challenge of distinguishing between the claims they accept and the claims they do not. Moral fictionalism is in the same boat as the error theorist about morality here; and, indeed, in the same boat as anyone who wants to be an error theorist about anything. It is more challenging to draw distinction between moral and non-moral concepts if one believes that many of our moral terms—eg courage, kindness, generosity—cannot be resolved into a purely factual component and a purely evaluative component (i.e. in the jargon, they are “thick” moral concepts). But even then it doesn’t seem to us that the challenge is unanswerable.

There is another concern about moral fictionalism, questioning the coherency of the project. The moral fictionalist takes herself to have good reason to believe that moral realism, while largely correct about the concept of morality and what moral features of the world would be like if there were any, is mistaken, since there are no positive moral truths. Assertions apparently committed to positive moral truths are useful to make, to communicate information about preferences, help our collective decision-making, or whatever moral talk is useful for. The moral fiction can do this because statements in it, along with non-moral claims imply other non-moral claims. To serve this purpose, there must be limits to what can be inferred from the moral fiction plus non-moral truths. The issue centres on whether these limits can be set.

Now, some of the error-theorists' arguments that the realist is wrong can be seen as arguments establishing their claims necessarily: one might think that nothing could even *possibly* have the sort of objective prescriptivity Mackie describes. If so, then the moral fiction is not even

possibly true. A less general worry, but still a serious one, is faced if the fictionalist embraces the supervenience of the moral on the non-moral. The combination of the moral fiction and the non-moral truths comprise a class of statements in which the non-moral is exactly as it is in this world, but in addition the class of statements contains many positive moral claims. However, the truth is that no positive moral claims are true. The problem is that if the moral supervenes on the non-moral, so that there cannot possibly be a moral difference without a non-moral difference, then the body of claims composed of our world's non-moral claims with the fiction's moral claims does not describe a possibility, given that the actual world, with the same non-moral truths and no positive moral truths, is a possibility (which it surely is).

The upshot from either concern is that the fiction, or the fiction together with the actual non-moral truths, describes an *impossibility*. This is of concern if the impossibility in question is a semantic or “broadly logical” impossibility, since it is standardly held that any proposition can be deduced from an impossibility. This would render the moral fiction useless for deriving non-moral implications, since one could derive anything at all that one liked (including the claim that we are pink elephants, our only desire is to go to the moon to eat treacle, or *anything*). If the fiction, or the fiction together with non-moral truth, is incoherent, then the fictionalist is in trouble.

The fictionalist has a range of responses. She could deny that there is anything semantically or logically impossible about there being positive moral truths and moral ontology of the sort the fiction claims: that it is a contingent matter of fact that all positive moral claims are false. (The plausibility of this will depend on the details of what is in the fiction). She could deny the supervenience of the moral on the non-moral, and say that it is perfectly possible that a world be just like ours non-morally, with the addition of a range of moral truths and moral ontology. However, denying the supervenience of the moral on the non-moral is taken to be unattractive by many, especially as it seems we are presented with moral features of situations only insofar as we are presented with non-moral features—in Gilbert Harman's example, we see the morally wrong act of deliberate cruelty *by* seeing a group of children setting a cat on fire.<sup>51</sup> There is also the thought that our concept of morality is not of something “free floating”: if there are

---

<sup>51</sup> Harman (1977), ch.1.

moral properties, objects have them because of their non-moral makeup, rather than having a moral “aura” descend upon them independently of how they are non-morally.

Another response is to accept a restricted supervenience claim, of one sort or another. Supervenience holding of less than logical necessity, or at its weakest mere intra-world supervenience, arguably captures at least some of our supervenience intuitions without denying the fictionalist the ability to assert the logical compossibility of the moral fiction and the non-moral facts. One attractive option is a “contingent supervenience claim”<sup>52</sup>. Just as tempting a claim is that there is no difference in the mental without a difference in the physical, *at least in worlds like ours*. In this world, as a physicalist takes it to be in any case, to have a mental state is a matter of being a physical structure with various informational and dispositional properties, or some neurological intrinsic features. However, a functionalist like David Lewis is willing to admit that ectoplasm or Cartesian spirit, or divine substance could have the right sorts of features too, in those worlds where such things can be found. Similarly, the supervenience of the moral on the natural might be taken to be a “contingent supervenience claim”—there is no difference in the moral without a difference in the natural in worlds at all like ours, but that is not to say they may not come apart in distant but still logically possible worlds where one can find, say, objective Moorean properties detected by intuition and which motivate when agents become aware of them.<sup>53</sup> We think that even realists should be tempted by this option—and fictionalists who avail themselves of this option have the advantage of supervenience, without rendering it logically impossible for there to be a world where the natural is as it actually is but the moral is as the story says it is.

Finally, the fictionalist may accommodate the supervenience intuition while denying supervenience. The fictionalist may say that the supervenience itself is only fictional: for instance, that it is a constraint on the fiction that what is to be true according to the fiction about moral matters is not to differ between worlds which agree in non-moral matters. This might be a natural way to go if the fictionalist takes positive moral claims to be false of semantic or logical necessity. But this no-difference-in-what-is-fictionally-true-about-morality-without-a-

---

<sup>52</sup> Lewis (1986), p. 16.

<sup>53</sup> Jackson (1992) disagrees. While he accepts contingent supervenience in the case of the mental, he rejects it in the case of the ethical.

difference-in-what-is-true-about-the-non-moral approach can also be of use to a fictionalist who does take the non-moral facts and the claims of the fiction to be compossible. For one thing that a fictionalist might think is that there is no non-moral difference between this world and a logically possible world where what is fiction here about morality is fact: for such a fictionalist, supervenience of the moral on the non-moral fails (since there is a non-moral duplicate of the actual world which is not a moral duplicate of it), but the “no difference in what is fictionally true about morality without a difference in what is true about the non-moral” condition is satisfied, since the non-actual world in question is a world in which, coincidentally, what is to be true according to the moral story happens to be moral fact as well. Such a fictionalist could hold they have salvaged what is needed from supervenience (supervenience-according-to-the-fiction), while preserving the compossibility of the fiction and the non-moral facts.

As well as these attempts to argue that the fiction of morality is consistent, even when it respects some supervenience constraint or other, the moral fictionalist has another alternative. The moral fictionalist could embrace what they take to be a necessarily false fiction about morality and its relation to the natural. The fictionalist following this strategy should deny that this would lead to triviality, for they would deny that inconsistent fictions are trivial: that anything whatever is true according to an impossible (or even inconsistent) fiction. There are many approaches to reasoning coherently about impossible situations available, and a moral fictionalist with suitable logical machinery can employ an impossible fiction without compromising their beliefs about the non-fictional world.

There is an obvious challenge for the moral fictionalist that we have so far said very little about. There are many possible fictional schemes a fictionalist might employ, which are different both in terms of what non-moral inferences are easy to draw from them from a given body of non-moral claims, and in terms of the style of moral theory, simplicity of functioning, familiarity, and so on. What fiction should a fictionalist recommend?

The answer to this question will, understandably, depend on many things, and will of course be answered differently by different possible moral fictionalist positions, with different justifications. Some preliminary remarks on our part, however, might prove helpful. One

thing that will make a great deal of difference is whether the moral fictionalist is a descriptive or prescriptive fictionalist: if the moral fictionalist holds that, as a matter of descriptive fact, we already engage in fictionalist behaviour, then it will be a matter of discovery which of the many possible fictions is the one that is actually employed on any given occasion. Such a discovery is presumably to be made by means similar to elucidating theories implicitly held by people in other areas of inquiry (with the difference that in this case the theory is not believed, but merely treated as if it is for some purposes). Examination of what people say about different moral problems, consulting of intuitions, perhaps the removal of the distorting effect of explicitly formulated meta-ethical hypotheses which mask people's pre-theoretic dispositions to make moral assertions, and so on, may all play a role.

Finally, fictionalism faces a challenge common to a wide range of non-realist approaches: how to do justice to the phenomenology of our ordinary moral experience—the nature of moral reasoning and deliberation, its connection with emotion and action, etc.<sup>54</sup> Most of us take ourselves to have a genuine concern to discover what is right and wrong and to do the right thing; and it does not seem to us, when we seek to arrive at answers to questions of what we should do, that we are simply striving to play the game properly or keeping up appearances. In short, our unreflective moral practices seem realist, not non-realist or fictionalist, in nature.

One obvious route of reply for the fictionalist is to explain away this phenomenology as the product of a pre-theoretic temptation towards moral realism, or at least as the product of our engagement in realist practices. Especially if the fictionalism being proposed is of the prescriptive variety, it may be neither surprising nor worrying that it may fail to account for some aspects of our unreflective moral practice.<sup>55</sup> After all, prescriptive fictionalists are explicitly recommending fictionalism as a useful *revision* to our current unreflective moral practice. The objection is more obviously to the point against non-realist approaches that are proposed as *descriptions* at some level of our ordinary moral practice (such as Blackburn's quasi-

---

<sup>54</sup>For a well-known statement of this challenge see Wiggins (1988).

<sup>55</sup>Unless the objection is that accommodating the phenomenology of ordinary moral experience is essential if we are to avoid just changing the subject. But this is controversial.

realism or non-cognitivist accounts that claim to be capturing the meaning of our moral terms).

The descriptive fictionalist can respond by taking moral deliberation to be deliberation about the content of a story<sup>56</sup>; and questioning whether the phenomenology of reasoning and deliberation in the moral case really *is* so different from the phenomenology of reasoning and deliberation about the content of a fiction. For example, consider our apparent concern to get the answers to moral questions right. Fictionalism leaves a place for this. We are not simply making the answers up, we are attempting to discover what the story says, and perhaps also what is entailed in a reasonably rule-governed way from what the story says. As anyone who has been to a bible studies class or ventured into a debate about a Shakespearean drama can attest, coming up with the right answer to questions about what the fiction says, and about what is entailed by what the fiction says, can engage people with the same kind of intellectual and emotional intensity as moral questions appear to do. People can, and often do, really care about getting the answers to these sorts of questions right. So there are at least these significant similarities in the phenomenology. One possible point of disanalogy is that moral reasoning and deliberation seem relevant to action in a very wide range of circumstances, whereas deliberation about fictional truths—for e.g. the culpability of Hamlet’s mother or whether Star Trek physics really makes any sense—doesn’t seem to be action-guiding to the same extent. But the fictionalist can point out that life-choices and actions *are* sometimes very strongly influenced by our reactions to paradigm fictions. Indeed, there is a whole institution of didactic fictionalising—cautionary tales and the like—which presupposes that people are capable of such responses. If people can be moved to support generous welfare policies by *Grapes of Wrath*, or to oppose genetic testing by reading *Brave New World* etc, how much more might they be concerned to “get things right” with respect to the story of goodness, justice and decency, to which they are deeply attached?

## References

Armstrong, David (1989) *A Combinatorial Theory of Possibility*. Cambridge: Cambridge University Press.

Ayer, A. J. (1932) *Language, Truth and Logic*. London: Penguin, ch. 6.

---

<sup>56</sup> Another option is to take moral deliberation to be pretend deliberation about objective matters of fact.

- Blackburn, Simon (1987) "Morals and Modals" in (eds.) C. Wright and G. Macdonald, *Fact, Science and Value: Essays in Honour of A. J. Ayer's Language Truth and Logic*. Oxford: Oxford University Press. Reprinted in Simon Blackburn (1993), *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- (1984) *Spreading The Word*. Oxford: Oxford University Press.
- (1993a) "Errors and the Phenomenology of Value", *Essays in Quasi-Realism*. New York: Oxford University Press, pp. 149–165.
- (1993b) "Attitudes and Contents", *Essays in Quasi-Realism*. New York: Oxford University Press, pp. 182–197.
- (1995) "Evaluations, Projections and Quasi-Realism" in Michael Smith (ed.), *Meta-ethics*. Aldershot: Dartmouth, 1995), pp. 3–38.
- Brink, D. (1989) *Moral Realism and the Foundations of Ethics*. Cambridge University Press: Cambridge.
- Bury, R.G. (1936) *Sextus Empiricus III*, Harvard University Press.
- Churchland, Paul (1988) *Matter and Consciousness*. Cambridge, Mass.: MIT Press.
- Field, Hartry (1980) *Science Without Numbers*. Oxford: Blackwell.
- (1991) *Realism, Mathematics and Modality*. Oxford: Blackwell.
- Crimmins, M. (1998) "Hesperus and Phosphorus: Sense, Pretense, and Reference". *Philosophical Review* 107.1: 1-47
- Currie, Greg (1990) *The Nature of Fiction*. Cambridge University Press: Cambridge.
- Garner, Richard (1993) "Are Convenient Fictions Harmful To Your Health?", *Philosophy East & West* 43: 87–106
- Geach, P. T. (1965) "Assertion", *Philosophical Review*, 74: 449–465
- Hare, R. M. (1952) *The Language of Morals*. Oxford: Oxford University Press.
- (1995) "Universal Prescriptivism" in Peter Singer (ed.), *A Companion to Ethics*. Oxford: Blackwell, ch. 40.
- Harman, Gilbert (1977) *The Nature of Morality*, Oxford University Press, New York.
- Hinckfuss, Ian (1987) *The Moral Society—Its Structure and Effects*. Number 16, Preprint Series In Environmental Philosophy, Australian National University.
- (1993) "Suppositions, Presuppositions, and Ontology". *Canadian Journal of Philosophy* 23.4: 595–617
- (1990) "Relevant Facts and Suppositions: A New Analysis of Conditionals". *Logique et Analyse* 131–132: 215–241
- Jackson, F. C. (1977) "Statements about Universals" *Mind* 76: 427–429
- (1994) 'Realism, truth and truth aptness', critical notice of Crispin Wright (1992), *Truth and Objectivity*. Cambridge, Mass: Harvard University Press in *Philosophical Books*, 35, 2: 162–169
- (1992) Critical Notice of Susan Hurley, *Natural Reasons*, *Australasian Journal of Philosophy*, 70: 475–488

- Jackson, Frank, Oppy, Graham and Smith, Michael. (1994) "Minimalism and Truth-Aptness". *Mind* 103.411: 287-302
- Joyce, Richard (2001) *The Myth of Morality*. Cambridge: Cambridge University Press.
- Lewis, D. K. (1978) "Truth in Fiction" *American Philosophical Quarterly* 15: 37-46
- (1986) *On the Plurality of Worlds*. Oxford: Blackwell.
- Mackie, John. (1977) *Ethics: Inventing Right and Wrong*. London: Penguin.
- Nietzsche, Friedrich. (1968) *The Will To Power* (trans. Walter Kauffmann and R.J. Hollingdale; edited by Walter Kauffmann). New York: Vintage Books, especially Book Two, Part II.
- Ogden, C.K. (1932) *Bentham's Theory of Fictions*. London: Kegan Paul, Trench, Trubner & Co.
- O'Leary-Hawthorne, John and Price, Huw. (1996) "How To Stand Up For Non-Cognitivists". *Australasian Journal of Philosophy* 74.2: 275-292
- Mandeville, Bernard (1924) *The Fable of the Bees*. Oxford: Clarendon Press.
- Pap, A. (1959) "Nominalism, Empiricism and Universals 1", *Philosophical Quarterly* 9: 330-340
- Pigden, Charles (1991) *Non-Naturalism versus Nihilism: Coursebook*, University of Otago.
- Plato (1993) *Republic*, trans. Robin Waterfield, Oxford: Oxford University Press.
- Putnam, Hilary (1972) *Philosophy of Logic*. London: Allen and Unwin.
- Quine, W. V. O. (1969) *Ontological Relativity and Other Essays*. London: Columbia University Press.
- Railton, P. (1986) "Moral Realism", *Philosophical Review*: 163-207.
- Rosen, Gideon (1990) "Modal Fictionalism", *Mind* 99: 327-354
- (1993) "A Problem for Fictionalism about Possible Worlds", *Analysis* 53: 71-81
- (1995) "Modal Fictionalism Fixed", *Analysis* 55: 67-73
- Schueler, G. F. (1988), "Modus Ponens and Moral Realism", *Ethics* 98: 492-500
- Searle, John R. (1969) *Speech Acts: an essay in the philosophy of language*. Cambridge: Cambridge University Press.
- (1979) *Expression and Meaning: Studies in the Theories of Speech Acts*. London: Cambridge University Press.
- Smart, J. J. C. and Bernard Williams (1973) *Utilitarianism: For and Against*. London: Cambridge University Press.
- Smith, M. (1994) *The Moral Problem*. Blackwell: Oxford.
- (2000) "Moral Realism" in Hugh LaFollette (ed.), *The Blackwell Guide to Ethical Theory*. Oxford: Blackwell., pp. 15-37.
- Stevenson, C. L. (1944) *Ethics and Language*. New Haven: Yale University Press.

- (1963) *Facts and Values*. New Haven: Yale University Press.
- Sturgeon, N. (1985) "Moral Explanations" in D. Copp and D. Zimmerman (eds.), *Morality, Reason and Truth*. Rowman and Allanheld, pp. 49-78.
- Vaihinger, H. (1935) *The philosophy of 'as if': a system of the theoretical, practical and religious fictions of mankind* (translated by C. K. Ogden). London: Routledge & Kegan Paul.
- Van Fraassen, Bas C. (1980) *The Scientific Image*. Oxford: Clarendon Press.
- Walton, Kendall (1990) *Mimesis as Make-Believe*. Cambridge: Harvard University Press.
- Wiggins, David (1988) "Truth, Invention and the Meaning of Life" in Geoffrey Sayre-McCord (ed.), *Essays on Moral Realism*. Ithaca: Cornell University Press, pp. 127-165.
- Yablo, Stephen. (2001) "Go Figure: A Path Through Fictionalism". *Midwest Studies in Philosophy* 25:72-102
- Zalta, Edward N. (1992) "Fictional Truth, Objects, and Characters", *Blackwell Companion to Metaphysics*, J. Kim and E. Sosa (eds.), Oxford: Blackwell.
- (2000) "The Road Between Pretense Theory and Object Theory" in A. Everett and T. Hofweber (eds.), *Empty Names, Fiction, and the Puzzles of Non-Existence*. Stanford: CSLI Publications, pp. 117-147



Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

NOLAN, DANIEL;Restall, Greg;West, Caroline

**Title:**

Moral Fictionalism

**Date:**

2002-10

**Citation:**

Nolan, Daniel and Restall, Greg and West, Caroline (2002) Moral Fictionalism.

**Persistent Link:**

<http://hdl.handle.net/11343/33700>