

**HERMENEUTIC MORAL FICTIONALISM AS AN ANTI-REALIST STRATEGY**  
**(Please cite the final version in *Philosophical Books* 49, January 2008)**

STACIE FRIEND

*Birkbeck College, London*

Fictionalism has become a standard, if controversial, anti-realist approach in various domains of metaphysics. The fictionalist approach to a discourse is to combine a ‘representational semantics’—one that takes statements in the discourse ‘at face value’, as capable of truth or falsity—with a denial that engaging in the discourse requires a commitment to its truth. So, for instance, the fictionalist about mathematics takes ‘nine is the number of planets’ to be true so long as nine is the number of planets, false otherwise; in this respect she parts company with nominalists who offer a non-standard semantics for such claims. At the same time, in uttering ‘nine is the number of planets’ the fictionalist neither believes nor asserts what she says. Rather than believing the proposition expressed by her words, she *accepts* that proposition; she treats it as “somehow good or interesting or useful independently of [its] truth value”.<sup>1</sup> Rather than asserting the proposition, she makes *as if* to assert it—in Gideon Rosen’s terminology, she *quasi-asserts* it—in order to do something else (the ‘something else’ differing from theory to theory).<sup>2</sup>

Fictionalist proposals are often divided into two categories: revolutionary fictionalism (RF) and hermeneutic fictionalism (HF).<sup>3</sup> Revolutionary fictionalists argue that we *ought* to adopt the fictionalist attitudes described above with respect to a particular discourse. For instance, at the end of *Ethics*,<sup>4</sup> after having argued that moral discourse is deeply in error, John Mackie does not propose that we abandon speaking of right and wrong; instead, he suggests that we continue to employ that discourse as a ‘useful fiction’. Similarly, the revolutionary fictionalist claims that although we might take discourse about numbers or possible worlds or moral facts (etc.) to be false, there are practical reasons to persist in talking *as if* there are numbers or possible worlds or moral facts (etc.). Hermeneutic fictionalists, by contrast, claim that we *already* adopt the fictionalist attitudes described above with respect to a given discourse. Hermeneutic fictionalism is widely accepted for at least some discourse about fictional characters. If we do not believe that Hamlet exists, we neither believe nor assert that Hamlet is Danish. In uttering ‘Hamlet is Danish,’ we quasi-assert that this is so—speaking as if there is a Hamlet—typically to get across something else (presumably something like: according to the fiction, Hamlet is Danish).

In keeping with the division between RF and HF, there are two kinds of moral fictionalism. Richard Joyce and Daniel Nolan, Greg Restall and Caroline West have argued in favour of revolutionary moral fictionalism (RMF).<sup>5</sup> RMF is motivated by an error theory about moral

1. Mark Eli Kalderon, *Moral Fictionalism* (Clarendon Press, 2005), p. 108. All citations in text are to this book.
2. See, e.g., Gideon Rosen ‘Modal Fictionalism’, *Mind* 99 (1990), pp. 327–54 and ‘What is Constructive Empiricism?’, *Philosophical Studies*, 74 (1994), pp. 143–78.
3. The hermeneutic/revolutionary distinction comes originally from John Burgess, in ‘Why I am Not a Nominalist’, *Notre Dame Journal of Formal Logic*, 24 (1983), pp. 93–105.
4. John Mackie, *Ethics: Inventing Right and Wrong* (Penguin, 1977).
5. Richard Joyce, *The Myth of Morality* (Cambridge University Press, 2001) and ‘Moral Fictionalism’, in M.E. Kalderon (ed.), *Fictionalism in Metaphysics* (Clarendon Press, 2005), pp. 287–313; Daniel Nolan, Greg Restall

discourse (and thought).<sup>6</sup> Assume that arguments against the reality of moral facts and moral properties are sound. And assume that we should take such statements as ‘Abortion is wrong’ at face value: as predicating wrongness of abortion. Then in believing or asserting that abortion is wrong, we believe or assert something false. The same applies to (most of) the rest of moral discourse. Having become convinced of this massive error, should we give up ordinary moral talk? According to RMF, the reduction of moral talk to some other less ontologically weighted discourse is infeasible, and its elimination inconvenient. Instead, the fictionalist should continue to deploy standard moral discourse as a *façon de parler*: like Berkeley, she may speak with the vulgar and think with the learned. The enlightened fictionalist merely accepts, and merely quasi-asserts, the proposition expressed by ‘Abortion is wrong’, perhaps to get across a claim about maximizing utility or prudential self-interest, or to express a non-cognitive attitude. This version of MF is revolutionary because it proposes a revision of our erroneous ordinary moral practice. The revision is not, however, in what we say or think—we are supposed to continue with our ordinary discourse—but in our attitudes toward what we say or think.

By contrast, hermeneutic moral fictionalism (HMF), as defended by Kalderon, is motivated not by metaphysical qualms but by epistemological considerations. According to HMF, we do not believe that abortion is wrong, we merely accept it; and in uttering ‘Abortion is wrong’ we do not assert that it is, we merely quasi-assert it. As a result we are not, *ipso facto* not erroneously, committed to a realm of moral facts and properties. (All of this is compatible with there being moral facts and properties; where RMF is atheistic about the moral facts, HMF is agnostic.) Assume that we have reason to adopt non-cognitivism about moral discourse. The standard non-cognitivist interpretation is that despite appearances, moral statements do not predicate moral properties and are not in any substantive sense capable of truth or falsity; rather, they express non-cognitive attitudes such as approval or disapproval. But this kind of non-factualist semantics is subject to numerous criticisms, most notably the Frege-Geach objection. Kalderon aims to avoid these criticisms by distinguishing non-cognitivism as a psychological thesis from non-factualism as a semantic thesis. HMF interprets ‘Abortion is wrong’ as true only if wrongness is a real property, false otherwise. On HMF, the truth or falsity of the statement does not matter, because in uttering ‘Abortion is wrong’ we merely quasi-assert it in order to do something else: in this case, to express non-cognitive moral acceptance (or the lack thereof).

It is typical to describe the difference between various forms of RF and HF as reducing to the distinction between *ought* and *do*: according to RF we ought to treat a domain of discourse in just the way that according to HF we already do. From this it would seem to follow that for a given domain, at most one of HF or RF is appropriate. If the discourse is already fictionalist, no fictionalist revolution should be required. In that case, assuming that one is already convinced that moral realism is unattractive, Kalderon’s HMF competes not only with standard non-cognitivist approaches (i.e., those that adopt a non-factualist semantics) but also with RMF. The question is why we should accept HMF over either of its anti-realist rivals. I shall sketch some

---

and Caroline West, ‘Moral Fictionalism versus the Rest’, *Australasian Journal of Philosophy*, 83 (2005), pp. 307–30.

6. I usually follow fictionalists in speaking only of ‘moral discourse’, but ‘moral discourse and thought’ should be understood.

reasons to doubt that HMF offers a viable alternative to either RMF or non-factualist non-cognitivism (NFNC).

RMF and HMF may be distinguished according to the place of error in each theory. According to RMF, our moral practices are in error: we think that there are moral facts when there are none. RMF counsels us to give up these erroneous beliefs and to take a different attitude toward moral claims. Because Kalderon denies that we believe or assert moral claims, he denies that our moral practice is in error. However, given that most people would characterize themselves as having moral beliefs and making moral assertions, Kalderon must claim that this characterization is mistaken. He says that “moral discourse is systematically misleading, for moral discourse is apparently cognitive. . . . After all, the practitioners conceive of themselves as cognizing moral facts—they unhesitatingly ascribe moral beliefs to one another” (p. 141). Thus Kalderon is committed to an error theory of our meta-epistemological and meta-linguistic practices: we think that we believe that there are moral properties and that we make moral assertions, when in fact we do neither.

From this perspective, HMF appears much less plausible than RMF. Kalderon agrees with the proponents of RMF that we believe that our moral discourse is cognitive and that we take ourselves to be committed to the moral facts. RMF takes these beliefs to be accurate; we really are engaged in a cognitive moral discourse that commits us to the moral facts. So RMF takes us to be wrong, not about our own explicitly held beliefs, but about the metaphysical facts, arguably a domain over which the layperson has little authority. By contrast, HMF denies that our beliefs about our own discourse and thought are accurate; this is a more surprising failure of first-person authority.<sup>7</sup> This analysis raises several questions. First, why does Kalderon think that our apparently cognitive moral practice is in fact non-cognitive? Second, how is it possible for us to be so massively in error concerning our own practice?

Kalderon offers two arguments in favour of non-cognitivism, only one of which will concern me here.<sup>8</sup> The *argument from intransigence* is roughly that if moral acceptance were cognitive, then in a disagreement about reasons in the course of moral inquiry we would be under a lax obligation to enquire further into the grounds of acceptance; but (Kalderon argues) we are not under such an obligation, so moral acceptance is non-cognitive. For example, someone convinced that abortion is morally permissible will not feel compelled to revisit his reasons or investigate his opponents’, even when confronted by a rational person who disagrees. This is not simply a claim about how people behave in moral inquiry. Kalderon’s argument is that if we reflect on our moral practices, we will recognize that they are not guided by cognitive norms. The most plausible way to think of this argument is in terms of longstanding moral debate. There comes a point in such a debate where opponents are likely to give up trying to convince each other, where reasons just seem to run out. Kalderon’s claim is that this can happen even when the

7. A similar point is made against other forms of HF by Jason Stanley, ‘Hermeneutic Fictionalism’, in P. French and H. Wettstein (eds.), *Midwest Studies 25: Figurative Language* (Blackwell, 2001), pp. 36–71. I think that the objection is more problematic for Kalderon’s proposal than for Stanley’s opponents.

8. Kalderon’s second argument, the *argument from aspect shift*, is that construing moral acceptance as non-cognitive best explains the distinctive phenomenology associated with accepting a moral claim. The role of this argument is primarily to give a positive account of the kind of non-cognitivism Kalderon has in mind, however; the argument from intransigence bears the weight of establishing non-cognitivism.

opponents take each other to be fully informed, fully rational interlocutors. He cites a comment by Hilary Putnam regarding his longstanding debate with Robert Nozick. Putnam claims that although he and Nozick “each have the highest regard for each other’s minds”, they each have “something akin to *contempt* . . . for a certain complex of emotions and judgements in the other” (quoted p. 36).

I do not think the argument for intransigence is convincing. It is compatible with Putnam’s claim—and his comment is cited by Kalderon only as an illustration—that he and Nozick each believe they are right about the moral facts. Where we have given up trying to convince our opponents, this need not be because we take further reasons to be without force; it could be because we take our opponents to be unable to see their force. Contrary to Kalderon’s assertion, we suspect that our interlocutors either are not fully informed or are not fully rational. Of course Kalderon’s is not the only argument against moral cognitivism. There are other more traditional arguments that seek to establish that our moral practices essentially involve expressions of non-cognitive attitudes, and many have found these arguments compelling.

Suppose, then, that we are already convinced that non-cognitivism is correct. The advantage of Kalderon’s proposal is that our commitment to non-cognitivism need not be coupled with a commitment to a non-factualist semantics. Traditional expressivist accounts of moral discourse run into Frege-Geach worries about embedding. To take an example Kalderon discusses (p. 57), consider the naïve expressivist who take utterances of (i) ‘Lying is wrong’, not to represent lying as having the property of wrongness, but instead to express disapproval of lying. The problem is that (ii) ‘If lying is wrong, then getting one’s little brother to lie’ need not express disapproval. The expressivist cannot say that the embedded and unembedded occurrences of ‘lying is wrong’ have different meanings, because then he would fail to explain the validity of an argument that derives from (i) and (ii) the following conclusion (iii) ‘Therefore, getting one’s little brother to lie is wrong’.

More sophisticated versions of NFNC, such as Blackburn’s quasi-realism and Gibbard’s norm-expressivism,<sup>9</sup> are designed precisely to avoid the Frege-Geach objection. Kalderon’s claim is that the complicated apparatuses proposed by these accounts are unnecessary if we just accept that the propositions expressed by our moral utterances are fully representational (though possibly false); they should be taken at face value. We must make a distinction between, on the one level, the *fictional content* of our utterances—the proposition we quasi-assert—and, on another level, what we are doing in quasi-asserting that content, which for Kalderon is expressing “an amalgam of cognitive and noncognitive attitudes” (p. 129). It is compatible with Kalderon’s view that the right analysis of this second level is given by one or another version of NFNC. So if non-cognitivists adopt his approach, they can say what they want to say while simply ignoring the Frege-Geach objection. This would seem an attractive option.

The option is attractive, however, only to the extent that we have an account of the fictional content of moral utterances. What exactly is the proposition expressed by an utterance of ‘Lying is wrong’? Like most fictionalists, Kalderon does not answer this question. He points out that

9. See especially Simon Blackburn, *Spreading the Word* (Oxford University Press, 1984) and *Essays in Quasi-Realism* (Oxford University Press, 1993); Alan Gibbard, *Wise Choices, Apt Feelings* (Harvard University Press, 1990). Blackburn argues that quasi-realism is not a version of fictionalism in ‘Quasi-Realism no Fictionalism’, in Kalderon (ed.), *Fictionalism in Metaphysics*, pp. 322–38.

realists typically do not answer the question either; the fictionalist merely claims that whatever the realist answer, the same answer can be used by fictionalism.<sup>10</sup> This assumption is problematic. It is open to the realist to hold that the semantic value of ‘wrong’ is a moral property. Even fictionalists who are merely agnostic rather than atheistic about the existence of moral properties cannot adopt this approach. The fictionalist might reply that such a semantics is implausible on independent grounds, for it seems to imply that any predicates that fail to pick out properties must be meaningless. Of course some fictionalists embrace this implication: they claim that a given class of utterances may fail to express any (complete) propositions at all. This is the line taken, for example, by Kendall Walton and other direct reference theorists with respect to statements about non-existents.<sup>11</sup>

Suppose, however, that we reject this approach and grant that ‘wrong’ is meaningful even if it does not pick out a property, so that ‘Lying is wrong’ expresses a proposition. In that case the anti-realist surely owes us some account of what this proposition looks like. Whatever account this may be, it is unlikely that it will take moral utterances ‘at face value’. This is obvious for discourse about fiction. Taken at face value, ‘Hamlet is a fictional character’ is a true statement about someone or something called ‘Hamlet’; for this reason the face value interpretation is often used to support realism. A realist might make a similar argument about ‘Lying is wrong’. So the fictionalist account of the semantics will have to depart in some way from the surface structure of such utterances. The right semantics might well turn out to be as complex as sophisticated versions of NFNC propose. The apparent advantage of HMF over NFNC should diminish accordingly.

In addition, although HMF is a logically possible position, I am not convinced that non-cognitivism can in any substantive sense be separated from non-factualism. To see why, we must return to the question of how we could be so radically mistaken about what we are doing when we think moral thoughts and make moral claims. Kalderon offers several replies, the most important of which is that the representational nature of moral discourse is likely to conceal the fact that morality is a fiction:

Mistaking making as if to believe for belief is facilitated by the fact that attributions of moral belief are true within the moral fiction. . . . That people believe the moral claims that they accept would be part of the extended moral fiction. Moreover, if attributions of moral belief are fictionally true, and moral pretense is unwitting, then it would be easy to mistake the fictional truth of such attributions for genuine truth, and so mistake a non-cognitive moral fiction for the cognition of the moral facts. (p. 156)

Some find it implausible that we could be engaged in the sort of unwitting pretence that various forms of HF postulate.<sup>12</sup> After all, when it comes to at least some discourse about fictions—the

10. This and other points in this paragraph were raised by Kalderon in conversation.

11. Kendall Walton, *Mimesis as Make-Believe: On the Foundations of the Representational Arts* (Harvard University Press, 1990), Ch. 10–11. See also Fred Adams, Gary Fuller and Robert Stecker, ‘The Semantics of Fictional Names’, *Pacific Philosophical Quarterly*, 78 (1997), pp. 128–48; David Braun, ‘Empty Names, Fictional Names, Mythical Names’, *Noûs* 39 (2005), pp. 596–631.

12. For instance, Jason Stanley, *op cit*.

domain that inspires other forms of fictionalism—our pretence is far from unwitting. In saying that Hamlet is Danish, there is a sense in which we are knowingly engaged in the pretence that there is a Hamlet, and we exploit this pretence in order to talk about the fiction. Thus Evans called this use of an empty name ‘conniving’ to contrast it with a case of error.<sup>13</sup>

Hermeneutic fictionalists often argue that we are in this position with respect to other domains of discourse. For instance, Stephen Yablo describes acceptance as *simulating belief*:

Someone is simulating belief that S if although things are in relevant respects *as if* they believed that S, when they reflect on the matter they find that they do not believe it; or at least are agnostic on the matter; or at least do not feel the propriety of their stance to depend on their belief that S if they have one.<sup>14</sup>

Acceptance as simulating belief does not require a deliberate effort in the way that paradigmatic games of make-believe with fictions do. A fictionalist interpretation of our acceptance that the sun rises and sets, or that things are in absolute motion or at absolute rest, is appropriate, not because we engage in explicit pretence, but because in our more reflective, literal-minded moments we would admit that we do not really believe any of this. Similarly, Yablo argues that we merely accept that there are numbers because we do not typically think it matters whether or not there are such things for arithmetical discourse to proceed apace. It does not matter to us whether or not arithmetical claims are literally true.

We might find this line of argument doubtful. Perhaps the literal truth of mathematics matters little to the layperson—who is after all likely to find the question of whether numbers *really* exist at best odd and at worst incomprehensible—but practising mathematicians (arguably) have convictions on this subject. The same concern has been raised about Bas van Fraassen’s constructive empiricism (CE).<sup>15</sup> According to van Fraassen, science aims at empirical adequacy rather than truth, and acceptance of a scientific theory involves the belief that it is empirically adequate. Yet scientists typically believe that their theories are true not only of the observable phenomena, but also of the posited underlying mechanisms. Van Fraassen denies, however, that CE is a thesis about the beliefs or intentions of individual scientists.<sup>16</sup> In engaging in science, the scientist pursues empirical adequacy even if, as an individual, she takes the theories to be true. The aim of science is a philosophical question rather than a sociological one. Does this mean that scientists are in error with respect to their own understanding of their work? Probably not, says van Fraassen: he argues that if we ask scientists the right questions, we will find that upon reflection, they understand science to aim at empirical adequacy.<sup>17</sup>

Given that Kalderon explicitly compares HMF to CE, we can ask whether he could make the same reply as van Fraassen. It seems unlikely that upon reflection we (the moral practitioners) would agree that our moral practices are indifferent to the existence of moral facts. On the other

13. Gareth Evans, *The Varieties of Reference* (Clarendon Press, 1982), Ch. 10.

14. Stephen Yablo, ‘Go Figure: A Path through Fictionalism’, in P. French and H. Wettstein (eds.), *Midwest Studies in Philosophy 25: Figurative Language* (Blackwell, 2001), p. 90.

15. See Bas C. van Fraassen, *The Scientific Image* (Oxford, 1980).

16. In ‘Gideon Rosen on Constructive Empiricism’, *Philosophical Studies*, 74 (1994), pp. 179–92.

17. *Ibid.*, pp. 187–88.

hand, presumably we are supposed to come to this conclusion upon reading Kalderon's book. Yet Kalderon does not think that our moral practice would remain unchanged if everyone came to recognize the fictionality of morality. One obvious reason, often cited against RMF, is that it is hard to see how a moral fiction could play the motivational role we ascribe to moral considerations. But Kalderon offers another reason. In the final part of the book he claims that intransigence is incompatible with moral authority, so that our moral discourse ought not to be the sort of fiction it is (pp. 174ff.). We should be open to the possibility that our current moral discourse fails to track authoritative reasons; one way this might be so is that there are moral facts that provide the relevant reasons, though this is not the only way. A renewed moral inquiry that rejects intransigence could take any number of forms, and Kalderon does not speculate on which form it is likely to take. The point, though, is that if everyone came to agree with Kalderon's argument, it is unlikely that our moral practices would remain the same. HMF as Kalderon presents it is unstable.

Part of the instability is merely a feature of the particular argument for non-cognitivism that Kalderon gives, namely the argument from intransigence.<sup>18</sup> But there seems to be instability inherent in any form of HMF. HMF requires that we use a kind of discourse suited to one sort of activity—expressing beliefs and assertions—in order to do something quite different. Why don't we use a language better suited to our aims? Kalderon's answer in the book seems to be that there is no better way to express moral acceptance than via the moral fiction, for two main reasons. First, it is possible that moral sensibilities might differ so that there will not be a single account of moral acceptance; in that case, we could not offer a uniform 'translation' of moral talk into non-moral. Second, it could be that "[i]n order to specify the relevant affect one needs to specify the relevant moral appearance, and in order to specify the relevant moral appearance one needs to specify what it is an appearance of—the moral property apparently instantiated by some aspect of the person's circumstance" (p. 134). These arguments are somewhat obscure; for example it is not clear how to understand a 'moral property apparently instantiated'. But more importantly, the claim that our moral discourse is suited to the expression of non-cognitive attitudes is the standard argument for NFNC. The proponent of NFNC will just take the reasons to think that our standard moral discourse is suited to expressing a certain affect and drawing attention to certain features of a situation as reasons to think that this discourse functions to do just that. Why not just accept this claim? Kalderon's answer is, of course, that NFNC faces difficulties in developing a non-factual semantics. But so too does HMF. Moreover, NFNC is not committed to the error theory and accompanying instability that characterize HMF.

Kalderon is thus in a difficult position. On the one hand, HMF requires that our moral practices be suited to the expression of non-cognitive attitudes: else why think that they are non-cognitivist? But, on the other hand, HMF requires that they must be unsuited to this expression: else why think there is a cognitivist fiction? It is only if we take discourse to be representational that we are led to think that it is cognitive; and it is only because we think that it is cognitive that we are in error. So it is essential for HMF to be an accurate description of our practice that we think we are engaged in cognitive arguments about reasons. If the discourse did not appear to be cognitive, we would not be taken in. Once we realise that we have been taken in, there is no

18. As Kalderon suggested in conversation.

reason to think that our discourse would continue as before, and good reasons to think that it should not.

Perhaps, though, Kalderon is wrong about what we would do if we discovered our error. According to RMF, if we realised that morality was a fiction we would nonetheless find it useful to retain our ordinary moral practices while adopting a different attitude toward them. Kalderon could take the same route. Assuming we can make sense of this option (which I find implausible), it does not help HMF, which would then seem to collapse into a version of RMF. The only difference between HMF and RMF would be with respect to where the error in our practice is located: in our beliefs about the moral facts, or in our beliefs about our beliefs about the moral facts. If we accept one or the other attribution of error, but find it too inconvenient or difficult to give up speaking as if there are moral facts, RMF says that we can continue to do so in a conniving fashion. But if this is so, it is hard to see why anti-realists should—apparently only temporarily—adopt HMF and its more implausible error theory. We have seen that reasons to think that our moral practices are non-cognitivist are reasons to accept NFNC. Now it seems that reasons to think that our moral practices are in error are reasons to accept RMF (or, if that option is infeasible, to revise our moral discourse). If we do not like either of these approaches, moral realism should be all the more attractive.<sup>19</sup>

19. I would like to thank Mark Kalderon for helpful discussion that clarified several aspects of his argument. Thanks also to Fraser MacBride for valuable discussion of related issues.