Morality, schmorality
Richard Joyce

In his contribution to this volume, Paul Bloomfield analyzes and attempts to answer the question "Why is it bad to be bad?" I too will use this question as my point of departure; in particular I want to approach the matter from the perspective of a moral error theorist. This discussion will preface one of the principal topics of this paper: the relationship between morality and self-interest. Again, my main goal is to clarify what the moral error theorist might say on this subject. Against this background, the final portion of this paper will be a discussion of moral fictionalism, defending it from some objections.

Bloomfield is correct to claim that the best way of removing the appearance of tautology or poor formation from the question "Why is it bad to be bad?" is to gloss it as elliptical for something along the lines of "Why does being morally bad have a deleterious effect on my self-interest?" The two "bad"s are intensionally non-identical: one (I will assume) refers to a non-moral notion of prudential badness (whatever is, all things considered, harmful to one's welfare[1]), while the other refers to a kind of ostensibly distinct *moral* badness. Though both notions have enough intuitive meat to them for discussion to proceed, neither is unproblematic. (I will return to these problems later.) On this interpretation, the question "Is it bad to be bad?" can be seen as an inquiry concerning whether two intensionally non-equivalent concepts are such that in fact (or even, perhaps, necessarily) the extension of one includes the extension of the other.[2] To ask the question "*Why* is it bad to be bad?" is to presuppose that this *is* the case, and to inquire in virtue of what this is so.

The moral error theorist thinks (i) that the predicate "…is morally bad" is a logical predicate (in contrast to the semantic noncognitivist, who thinks that it is a predicate only in a grammatical sense), (ii) that sentences of the form "ϕ is morally bad" are generally uttered with assertoric force (in contrast to the pragmatic noncognitivist, who thinks that such sentences are used to perform some other linguistic function), and (iii) that the predicate "…is morally bad" has an empty extension (in contrast to, e.g., the moral realist, who thinks that the property of moral badness is instantiated).[3] The third

---

1. Contra W.D. Falk (this volume), in this paper I am not using the term "prudence" to denote to a policy of risk-avoidance, but rather am identifying prudence with acting in whatever way advances one's interests, all things considered. (And the relevant notion of "interests" I am leaving unspecified.) I am happy also to use the term "expediency" for the same. Despite Falk's insistence that "expediency" must implicate some notion of *convenience*, my dictionary tells me that it also means simply "self-serving."
2. I take it that nobody will claim that the two concepts are co-extensive: that every act of imprudence is a moral wrong. My having a cup of coffee before going to bed may be prudentially foolish, but surely doesn't count as even a mild moral crime.
3. The options mentioned in this sentence are not intended to be exhaustive.

contention is the most controversial, and there are various reasons that might lead one to endorse it; it is not my intention in this paper to attempt to make any of these reasons compelling. Perhaps the error theorist thinks that for something to be morally bad would imply or presuppose that human actions enjoy a kind of unrestricted *autonomy*, while thinking that in fact the universe supplies no such autonomy (Haji 1998, 2003). Perhaps she thinks that for something to be morally bad would imply or presuppose a kind of inescapable, authoritative imperative against pursuing that thing, while thinking that in fact the universe supplies no such imperatives (Mackie 1977; Joyce 2001). Perhaps she thinks that for something to be morally bad would imply or presuppose that human moral attitudes manifest a kind of uniformity, while thinking that in fact attitudes do not converge (Burgess 1978). Perhaps she thinks that there exists no phenomenon whose explanation requires that the property of moral badness be instantiated, while thinking that explanatory redundancy is good ground for disbelief (Hinckfuss 1987). Perhaps she thinks that tracing the history of the concept *moral badness* back to its origins reveals a basis in supernatural and magical forces and bonds—a defective metaphysical framework outside which the concept makes no sense (Hägerström 1953). Perhaps she thinks all of these things and more besides.[4] The details are not important here; the point is that the error theorist accuses morality of being fatally flawed, such that any value system with the flawed element(s) extirpated simply wouldn't deserve the name "morality." The only detail that need be noted here about the moral error theoretic position is that it is usually restricted to the moral realm. Of course, in principle one could endorse a radical *global* error theory, in which case one would by implication be an error theorist about morality (along with modality, colors, other minds, cats and dogs, etc.), or one could be an error theorist about all normative phenomena, which, again, would include an error theory for morality. But typically the moral error theorist thinks that there is something *especially* problematic about morality, and does not harbor the same doubts about normativity in general. The moral error theorist usually allows that we can still deliberate about how to act, she thinks that we can still make sense of actions harming or advancing our own welfare (and others' welfare), and thus she thinks that we can continue to make sense of prudential "ought"s.[5] She allows that prudential badness is instantiated but insists that moral badness is not. Thus, on the assumption that the question "Is it bad to be bad?" amounts to an inquiry about the truth value of a universal conditional ("Is it the case that: for any *x*, if *x* is morally bad, then *x* is prudentially bad?"), the moral error theorist will think that the answer to the question is vacuously "Yes" because the conditional has a

---

4. For the sake of brevity I will talk as if the error theorist thinks there is only one thing problematic about morality. But of course an error theorist may be impressed by a number of considerations against morality. Perhaps morality has a lot of little or medium-sized problems—none of which by itself would ground an error theory, but all of which together constitute A Big Problem.

5. In this paper I assume that prudence naturally takes the form of a normative system, that it involves "ought" claims, reasons for action, etc. In fact one could deny this. All that is minimally necessary for believing in prudence is to accept that individuals can be harmed. Thus even if one thought that all "ought" claims are false—even all non-moral ones—one could still uphold that "…is prudentially bad" has a non-empty extension.

false antecedent irrespective of how the variable is instantiated.[6] (Note that she will also, for the same reason, answer "Is it good to be bad?" in the positive.) Thus, she will object to the presupposition behind the question "Why is it bad to be bad?" In this respect the question is, for her, not unlike "Why is it bad to annoy a witch?" Her answer is: "But you *can't* annoy a witch—there aren't any!"

But there is another way of understanding the elliptical element of the question that allows the possibility of the error theorist giving a substantive and interesting "Yes." If she treats the reference to moral badness as denoting the extension *that it is widely assumed to have*, the extension is not empty at all (though see below). After all, the error theorist is well aware that there is a broad range of actions—both types and tokens—that are widely thought to be morally bad: breaking promises, stealing, unprovoked violence, Hitler's Final Solution, gluttony, sloth, envy, etc., etc.[7] She can understand the question "Is it bad to be bad?" as "Will performing *these* actions [gesturing to those actions that are widely considered to be morally bad] have a deleterious effect on the interests of the perpetrator?" For token actions that have already been performed—e.g., Ernie's lying to Bert last week—the question must be either "Did this action have a deleterious effect on Ernie's interests?" or "Would performing an action of the same type have a deleterious effect on the interests of the perpetrator?" These are all questions that the error theorist might answer positively, thus allowing that the question "*Why* is it bad for me to pursue such things?" must have an answer.

(By comparison, suppose an anthropologist were studying a culture in which certain persons are considered to be witches. The anthropologist might recognize that it's a good thing—good for his research, that is—if he stays on friendly terms with these persons, even though he doesn't believe that they possess the supernatural powers necessary for actually being witches at all. He might say "It is good to be friendly to those persons that are hereabouts considered to be witches," but there would be nothing impermissible, or, in general, misleading, if he were, for convenience, to express this elliptically as "It is good to be friendly to the witches hereabouts.")

One problem with this interpretation is that there may be significant disagreement among the people "hereabouts" as to what counts as morally bad, such that even the predicate "… is widely assumed to be morally bad" threatens to turn up empty. After all, moral discourse, it is often observed, is characterized by a high degree of intractable disagreement. Perhaps, though, there are at least *some* things for which there is sufficient concurrence that we can speak of "what is widely assumed to be bad" (strangling babies?), and perhaps the error theorist confines her question merely to these actions. Or perhaps the error theorist just passes the buck to her interlocutor, and says: "Tell me what things you consider to be morally bad, and I will tell you whether (and, if so, in virtue of what) their pursuit is imprudent."

---

6. There are some complications here concerning (A) whether the domain of the variable is restricted to actual entities, and (B) whether the error theorist holds that moral predicates are necessarily empty or just actually empty. Addressing these complications is unnecessary.

7. Although for the sake of simplicity I tend to speak just of actions being morally bad, I don't mean to exclude morally bad character traits, states of affairs, intentions, policies, properties, objects, etc.

Another feature of the question to which attention should be drawn is the fact that it may receive different answers for different people, or for the same person at different times, or for the same person (or counterparts, if you prefer) at different possible worlds. Perhaps it will frustrate Ernie's interests to lie, but it won't frustrate Bert's interests to lie. Perhaps it will frustrate Ernie's interests to lie today, but he'll be okay if he waits till next Friday. Or perhaps it will frustrate both Ernie's and Bert's interests to lie, but it will do so for very different reasons: e.g., Ernie would have to live with crippling guilt, whereas Bert would be sent to bed without any dinner. Or perhaps as a matter of fact everyone has a prudential reason to avoid badness (and perhaps they all have the same reason), but there are possible circumstances where the pursuit of the bad would become prudentially good (for at least some persons).

It has been a long-standing aspiration of a certain school of moral philosophy—upon whose roll appears the name *Bloomfield, P.*—that all such contingent messiness could be swept aside by the provision of a universal, permanent, monolithic and (perhaps) necessary positive answer to the question. Bloomfield's solution is that all bad human agents undermine their self-respect and thus frustrate their own interests. As far as go the principal theses of this paper, Bloomfield may be entirely right. But I happen to doubt that he is, and I find my sense of courtesy to the good editor of this volume prevailed over by an intellectual urge to join the fray; hence I cannot forego making a couple of critical comments.

First, it should be noted that at best his argument shows that there is something self-damaging about a certain kind of radical, ubiquitous, all-encompassing, self-conscious attitude towards what is (widely assumed to be) bad: pleonexia. But whether there even *are* any such awful characters around is a moot question. The agents who perform those actions widely thought of as bad— breaking promises, stealing, etc.—are rarely inclined to appeal to Thrasymachean or Machiavellian iconoclasm to attempt to justify themselves. Most everyday wrongdoers,[8] I submit, believe that what they are doing *isn't* really bad (and that if others disagree it's because they're not properly acquainted with the details of the case). Wrongdoing is born of negligence as often as it is born of arrogance. Many wrongdoers castigate themselves for their actions, and even perform them regretfully. Wrongdoers are not always selfishly motivated by gewgaws: Consider a mafia hit man acting out of obligation and loyalty (perhaps even love) for the paterfamilias. Few wrongdoers fail to distinguish between the out-group (a domain of potential victims) and the in-group (a domain of friends, family, loved ones, those with whom one has binding obligations, etc.). In short, the pleonectic may be a fascinating philosophical case study, but he hardly represents the typical or paradigmatic instance of badness.

Bloomfield seems to think that the pleonectic represents the toughest case, and thus that if even Thrasymachus and his ilk can be shown to be harming themselves then surely

---

8. Here I am using the term "wrongdoers" in a purely descriptive manner: to pick out those people that are widely considered to be wrongdoers. Not wanting to beg the question against the error theorist, I should really keep the term in scare-quotes throughout, but I refrain from doing so for stylistic reasons.

all those more mundane wrongdoers must proportionally follow suit. But this expectation is, in my opinion, ill-founded. Though the pleonectic does in some sense occupy an extreme wing of villainy, it doesn't follow that any injury he does himself must by implication be suffered to a lesser degree by less radical wrongdoers. Consider, for example, the claim that the pleonectic lives a life without "true love" and lacking "real friendship." We can all accept that any human who chooses such an existence is very probably damaging himself. But what about a lesser wrongdoer who, say, is creative with the truth when filing his taxes or is needlessly curt to a taxi driver? It might be claimed that this person has harmed himself to a lesser degree by missing an opportunity for *some* true love and *some* real friends (i.e., the love and friendship of the victims?). But there is surely nothing wrong with this kind of loss *per se*, for *everyone*—even the thoroughly virtuous—must eschew *some* potential friendships. (I don't recall that Mother Teresa ever sent me a Christmas card.) I see no grounds for assuming that a mundane wrongdoer cannot enjoy the full complement of genuine friends, or that the occasional bit of everyday misconduct (directed at non-friends) must, to some small degree, undermine those friendships. Consider instead the claim that certain pleonectics must be guilty of psychological compartmentalization. We can all agree that extreme compartmentalization of one's thoughts and desires is a harmful state. But what about *a little* compartmentalization? There is presumably nothing wrong with "a bit" of compartmentalization, since, again, it is an attribute that *every* human exhibits; it's the nature of human psychology. It might be complained that in this context the term "compartmentalization" is intended to denote only the pernicious, pathological variety. But then we are free to deny that the mild transgressor must manifest *any* such attribute, and any insistence that he does so simply begs the question.

If I am correct that the harm that the pleonectic (allegedly) does himself derives from aspects of the very extremism of his attitude, then there are no grounds for thinking that a lesser degree of the same kind of self-harm is in store for the everyday moral transgressor. And thus we have not been shown how moral badness *per se* is self-injurious, but rather only how a proper subset of moral badness is bad—and a very small (and perhaps actually empty) proper subset of moral badness at that.

The second critical comment I will make against Bloomfield's argument is that it at best shows that there is some kind of *fault* with the pleonectic, but it not clear how this fault translates into an injury. The pleonectic, according to Bloomfield, has but a simulacrum of self-respect; what she takes to be self-respect is "faulty in its foundations." Because the pleonectic accords others no respect, she cannot coherently respect herself, for to do so would be based on the (allegedly) impermissible distinction that "I deserve more because I am me." Though the pleonectic may be quite convinced that he does have self-respect, he is in fact self-deceived. To grant Bloomfield this case (something that I am in fact very far from doing) would be to acknowledge that a milestone in philosophy has been achieved. An argument demonstrating the irrationality of wrongdoers is something that Simon Blackburn has described as the "holy grail of moral philosophy" (1984: 222). Nevertheless, even if Bloomfield's argument delivered the grail into our hands, this would not achieve the goal he set himself, for it is simply not clear how being

irrational or self-deceived entails doing oneself harm. "I am special because I am me" may be a misguided or irrational thought (though even this I am highly doubtful of), but why *self-harmful*?

If someone is habitually irrational in all her deliberations then it is not unreasonable to suppose that this will land her in various kinds of trouble; and it is not hard to see that self-deception will often be self-injurious. But to show that irrationality and self-deception are on very many occasions harmful is insufficient to establish that there is anything harmful about these phenomena *per se*. This is especially evident when the charge of irrationality/self-deception is so unobvious that it takes a philosopher to establish it—against a background of over two thousand years of like endeavors meeting with a body of staunch academic opposition. When the accusation concerns so inconspicuous and subtle a phenomenon, any assumption that the typical harms that issue from canonical and obvious irrationality/self-deception must also issue from the inconspicuous instances must be suspended. In other words, if Bloomfield were to succeed in demonstrating that every moral wrongdoer is to some extent self-deceived, then he would have shown us that the domain of self-deception is very different than it is widely assumed to be, and thus any previous assumptions about the general harmfulness of self-deception (based, as they are, on a different class of prototypes) would stand in need of reexamination.

Just as space allowed Bloomfield to make his case but briefly, so too I will not attempt to respond to his final "five things that could be said to Thrasymachus" in any detail. My main suspicion is that they are indeed things that could be said *to Thrasymachus* (i.e., to the pleonectic), but have considerably less force against a more everyday wrongdoer. That someone who cheats slightly on his taxes, or is needlessly discourteous to the taxi driver, is suffering from schizophrenia, that he must endure the anxieties of dissimulation, that he is missing the "the joy of seeing things as they actually are"—that he is leaning towards any of these wretched states *even slightly*—are, at best, optimistic claims in need of empirical support. Of course there is a kind of satisfaction that comes from a job done with moral integrity; but there is also a satisfaction that comes from getting away with something. Of the people who have experienced both, of course there are some who prefer the first kind of pleasure; but there are also, I'll wager, some who prefer the second kind. (Many of us are not insensitive to both kinds.) The latter people may very well be self-deceived—it is not my intention here to deny it—but what needs to be asked is whether they are *harming themselves*. To appeal to a "joy" that comes from having true beliefs may sound appealing—especially to a philosopher—but I don't think it stands up to scrutiny. Do true beliefs *always* bring this joy? I don't recall the last time I felt even a hint of ecstasy when contemplating that $1 + 1 = 2$. Perhaps Bloomfield means to restrict his comment just to a certain domain of epistemic success: a joy that comes from having true beliefs about our own value in comparison to that of other humans. Again, Bloomfield's opponent need not deny the very possibility of such a joy, nor even deny that it might be quite widespread. All she need deny is the universal claim that Bloomfield's argument requires if it is to succeed: that such joy is available to *anyone* in *any* circumstances, and that it can never be outweighed by a countervailing joy

that flows from gaining benefits (and not necessarily mere gewgaws) secured through an act of moral transgression.

It is not my intention to criticize Bloomfield's argument beyond these gestures, because the main point to which I want to draw attention is that as far as the moral error theorist is concerned Bloomfield could be 100% correct. Chances are, what the moral error theorist is likely to say in response to the question "Is it bad to be morally bad?" (understood as outlined above) is "Sometimes it is, sometimes it isn't." But were she instead to answer "Yes: always, for everyone, necessarily"—and then go on to justify this answer by appeal to Bloomfield's argument—she would in no sense undermine her commitment to a moral error theory. Embracing a moral error theory rationally eliminates from one's serious practical deliberations certain kinds of justification: One can no longer, for example, refrain from doing something because one believes that it is morally forbidden. But it implies nothing about what actions one should actually perform (or refrain from performing). Contrary to popular belief, the moral error theorist is not a scheming villain, acting pleasantly solely in order to avoid punishment or to lull her victims into complacency. (As Richard Garner puts it: "The amoralist need not be an immoral, heartless, selfish jerk who denies the obvious" (1994: 279).) The moral error theorist may have as much compassion, love, and generosity as anyone else; she will just not believe these characteristics, or their attendant actions, to be *morally* desirable.[9] Nor does the embrace of a moral error theory obviously exclude any particular *non-moral* forms of justification from figuring in one's deliberations. The moral error theorist may be motivated largely by compassion, or by self-interest, or by a sense of loyalty to her friends and family, or (more likely) a mixture of these things (and others besides) depending on the situation.[10] There is simply no reason to assume that having such a (non-moral) basis to one's deliberations is going to end up prescribing sneaky nastiness. On the contrary, for most people, in most ordinary situations, it is fair to assume that a proper sensitivity to such non-moral considerations is likely to favor acting in accordance with (what most people think of as) moral requirements. So the moral error theorist is as willing and able as anyone else to endorse claims such as "I ought not break promises," "I ought not steal," etc.—it is just that for her the "ought" is a non-moral one. And, as I say, perhaps the moral error theorist will read Bloomfield's paper and believe it, thus arming herself with a foundation for thinking that self-interest will *always and for everyone* come out on the side of morality. None of this jars her commitment to a moral error theory in the least.

If any of this seems uncomfortable, then it may be useful to consider an analogy. Picture a theistic error theorist—better known as an "atheist." Suppose there were a kind

---

9. We mustn't be distracted by the fact that such emotions as love and generosity are often called "*moral* emotions." If they warrant this label it is in virtue of the fact that they are considered morally praiseworthy, but it is clear that one can have these emotions without making any moral judgment. The moral error theorist does not have her position undermined if *others* choose to judge her character, actions, and emotions in moral terms.

10. There is, of course, a kind of loyalty that is based on judgments of moral obligation. I submit, however, that one can also have "feelings" of loyalty—feelings of attachment and affection that involve desires to protect the welfare of another person—that need not be "moralized" by the subject in the least.

of prescription that could be marked as "…according to God": "You ought not kill, according to God," "You ought not testify as a false witness against your neighbor, according to God," etc. The atheist is unmoved by these prescriptions *qua* divine commands; he doesn't believe in God, so doesn't believe that there are any commands issuing from God, so doesn't believe that one ought not kill, according to God. It hardly follows, however, that the atheist is inclined to go around killing, or, indeed, that his reluctance to kill is in any flimsier than that of the Pope. The atheist may be as determined to refrain from killing as anyone, for any number of reasons. Perhaps he thinks that it is morally wrong, perhaps he has so much sympathy for his fellow human that the thought sickens him, perhaps he recognizes certain forms of self-harm that would ensue from killing, perhaps all of the above. The atheist is still inclined to enthusiastically assert "I ought not kill"—and perhaps takes himself to have grounds for holding that this is true always and for everyone—but he will remain clear in his own mind that he is not employing the "ought…according to God" locution. And this, clearly, doesn't undermine his atheism in the least.

The comparison between atheism and moral error theory is useful to bear in mind when it comes to responding to a possible objection to what has been argued. The objection runs as follows:

> You error theorists argue that morality is flawed, yet you still think that we ought to refrain from stealing, keep promises, not initiate violence, and so on. But if the foundational moral question is "How ought one to live?" and you have answered this, by reference to self-interest, in such a way that the answer is "Keep promises, refrain from stealing, don't initiate violence, etc." then you *have* endorsed a morality. You have allowed that moral normativity can be identified with prudence (or at least with a proper subset of prudence). So your moral error theory collapses.

In order to understand the moral error theorist's response to this objection, first some distinctions must be drawn. We must note, to begin with, the sense in which even the moral error theorist "believes in morality": She believes that *moralities* exist, in the same way as the atheist recognizes that religions exist.[11] What the error theorist does not do is epistemically endorse any morality. I say "*epistemically* endorse" so as to exclude certain pragmatic ways in which a morality might be "endorsed," such as approving of its practical output (agreeing that one ought not break promises, ought not steal, etc.), or acknowledging that the institutions of morality are instrumentally beneficial.

Note also that the error theorist need not have granted that there *is* a systematic answer to the question "How ought one to live?" Perhaps Ernie should live one way, given his circumstances and upbringing, and Bert should live another, given his. (Indeed, it may be precisely in virtue of thinking that there is no answer to the question "How ought *one* to live?" that someone is a moral error theorist.) But let us suppose that we are

---

11. I'll assume without argument that endorsing a religion entails endorsing theism. Some might object to this (raising the case of Buddhism, for example), but the niceties of that debate do not interest me.

dealing with a kind of error theorist who, for whatever reason (perhaps having been convinced by Bloomfield's argument), accepts that the question can receive some kind of universal, systematic answer—that there is a way that "one" ought to live.

The above objection in fact suggests two challenges for this type of moral error theorist. The first is that acknowledging that the question "How ought one to live?" can receive any positive answer in itself constitutes or implies the epistemic endorsement of a morality. The second is that answering this question *in a way that underwrites a particular content* (not breaking promises, not stealing, etc.) constitutes the epistemic endorsement of a morality. In both cases the moral error theorist will offer much the same answer: She will disagree because she believes that there is something special about moral normativity (something that, she thinks, is deeply flawed) such that merely to answer how one ought to live, or even to answer it in a way that underwrites keeping promises, etc., is insufficient to amount to the epistemic endorsement of a morality. Imagine, by analogy, the atheist facing the objection that in so far as he thinks that there is a way we ought to live then he is, despite himself, really a theist, *because that's all there is to theism*. He will, quite rightly, object that that's *not* all there is to theism, that to epistemically endorse a theistic framework requires subscribing to some substantive metaphysical theses about the existence of a divine being who enjoys such properties as omnipotence, omniscience, etc. It is in virtue of disbelieving these theism-constituting theses that the atheist is an atheist. Similarly, the moral error theorist also thinks that to endorse any moral system requires subscribing to some substantive (and, presumably, "metaphysical," in some broad sense of the word) theses, and it is in virtue of her disbelief in these theses that she is a moral error theorist. (The kind of theses in question were pointed to in the third paragraph of this paper.)

In fact, the idea that giving *any* positive answer to the question "How ought one to live?" constitutes the epistemic endorsement of a morality seems highly implausible. Suppose the answer comes back: "Do whatever the hell you feel like." There would seem to be something terribly misleading in the insistence that living according to this rule constitutes endorsing a *morality*. (If one really wanted to stretch the word "morality" to this extent, the moral error theorist can always just disambiguate: "Well, okay, in that unnaturally strained sense of 'moral,' of course I endorse morality—but nevertheless there is a far more familiar customary usage regarding which I remain a disbeliever." We can imagine the atheist saying something comparable if faced with the serious assertion that *God is love*.) The objection, as it is stated above, contains an element that implies that not just *any* positive answer to this question will constitute the endorsement of a morality; rather, there appears to be a contentful constraint on what can count: Prudence (or a proper subset thereof[12]) becomes a candidate for constituting a morality only to the extent that it endorses keeping promises, refraining from stealing, not initiating violence, and so on. But even with the addition of this constraint on content, the moral error theorist will—for the same reason as before—remain unimpressed with the proposal that she has, despite herself, endorsed a morality. Whatever argument or arguments have led

---

12. See note 2. For the sake of brevity I will drop this qualification about proper subsets.

her to embrace moral skepticism will almost certainly constitute grounds for resisting this objection. To repeat: The moral error theorist believes that for something to be morally bad (say) would require the instantiation of some property that (i) is not supplied by the universe (as a matter of fact or necessarily), and (ii) is *essential* to moral badness, such that anything lacking this feature just won't count as moral badness. (For ease of reference, let us call this property the "special feature" that the error theorist attributes to morality.) Assuming that we are dealing with an error theorist who allows that there is nothing particularly fishy about prudence, then we are *ex hypothesi* dealing with someone who thinks that prudential normativity lacks the special feature that dooms moral normativity. Thus the moral error theorist will not think that prudence is a good contender for being identified with moral normativity: Someone whose deliberations are guided solely by prudential considerations—even if these considerations speak in favor of all the things that morality is typically assumed to prescribe—is not thereby epistemically endorsing a morality.

It might be objected—by a moral noncognitivist, for example—that morality is not the kind of thing that requires *epistemic* endorsement at all. It might be objected that the only kind of endorsement needed is practical, and that so long as a person is generally behaving himself then he is endorsing morality in the only sense that matters. However, the dialectical point that I am making is that whatever argument(s) have led a person to defend a moral error theory will include grounds for thinking that moral judgment is a matter of *belief*, that *epistemic* endorsement is coherent and called for. The objection under consideration is that the moral error theorist somehow undermines her own position if she accepts prudential normativity and accepts that it speaks in favor of general niceness. This objection cannot be founded on an insistence that noncognitivism is true, for the error theorist *ex hypothesi* won't agree to this.

For all that, noncognitivism could be true; nothing I say in this paper is designed to convince anyone otherwise. At no point is my intention to establish that the moral error theorist is *correct*. Perhaps the special feature that the error theorist attributes to morality is instantiated by the universe after all. Or perhaps the error theorist is mistaken in thinking that this feature is an *essential* characteristic of moral normativity; perhaps a kind of normativity lacking this feature would nevertheless satisfy enough of our other desiderata to count as the real thing. A moral philosopher advocating an error theory must be prepared to defend herself on both fronts. This job is made difficult by the fact that it is often extremely difficult to articulate precisely what it is that is so troubling about morality. And this failure need not be due to a lack of clear thinking or imagination on the error theorist's part, for the thing that is troubling her may be that there is something deeply mysterious about morality. The moral error theorist may, for example, perceive that moral imperatives are imbued with a kind of mystical practical authority—a quality that, being mysterious, of course *cannot* be articulated in terms satisfactory to an analytic philosopher. Such an error theorist is forced to fall back on vague metaphors in presenting her case: Moral properties have a "to-be-pursuedness" to them (Mackie 1977: 40), moral facts would require that "the universe takes sides" (Burgess 1978), moral believers are committed to "demands as real as trees and as authoritative as orders from

headquarters" (Garner 1994: 61), the phenomenology of believing oneself morally required to act is to think "Well, I just *have* to" (Joyce 2001: 141), and so on. Indeed, it may be the very perniciously vague, equivocal, quasi-mystical, and/or ineliminably metaphorical imponderabilia of moral discourse that troubles the error theorist.[13] (For useful discussion of this point, see Hussain 2004.)

As I have indicated earlier in this paper, it is not my intention on this occasion to present any particular error theoretic argument regarding morality. For a start, doing so would take too long, and, moreover, it is more useful here to keep things broad so as to give consideration to the moral error theorist in a generic sense (hence these unsatisfying references to a "special feature" that the error theorist attributes to morality). It might be thought that without presenting any particular argument it will be impossible to assess whether the error theorist is reasonable in claiming that prudential normativity cannot be identical to moral normativity. It might be thought that we really need to have the error theorist spell out what she takes the essential and problematic feature of morality to be, so we can judge whether she is correct in claiming that prudence lacks it. But in fact I think that we can get a pretty good taste of how that argument will go without committing our (usefully generic) error theorist to any particular line of reasoning. Indeed, it seems to me that anybody—whether error theorist or not—should be extremely uncomfortable about any proposal to identify moral imperatives and values with prudential imperatives and values.

Let us begin by thinking about how prudential normativity works. Suppose it is claimed "Ernie ought not eat cookies in bed," using a plain and simple prudential "ought." The sentence is true (with the prudential "ought") only if eating cookies in bed will harm Ernie in some way. Perhaps doing so will lead to crumbs in his pajamas, leading to sleeplessness. But it is possible that there is harm to other parties involved too. Perhaps what is under consideration is Ernie's decision to eat cookies *in Bert's* bed, thereby annoying (harming) Bert, which will lead to Bert retaliating against (harming) Ernie in some way. (Or perhaps God punishes Ernie, or perhaps Ernie pollutes his own soul, or perhaps Ernie fails to respect himself, etc.—the details don't matter.) The important thing to notice about a prudential "ought" that involves harm to more than one party is the counterfactual asymmetry between the harms: If in eating cookies in Bert's bed Ernie will harm himself but somehow (magically, perhaps) Bert will escape harm, then the prudential claim would remain true; but if in eating cookies in Bert's bed Ernie will harm Bert but will somehow manage to avoid the self-harm, then the prudential claim would have to be retracted. (In the latter case, of course, it may remain true that Ernie ought not to eat cookies in Bert's bed, *using some other kind of "ought."*)

---

13. Cf. Wittgenstein, who concluded that moral language is "nonsense" on the basis of his observation that moral discourse consists largely of similes, yet "a simile must be a simile for something … [but] as soon as we try to drop the simile and simply state the facts which stand behind it, we find there are no such facts" (1965: 10). Interestingly (in light of what I will discuss later in this paper), although he concludes that nonsense is "the very essence" of moral expressions (11), Wittgenstein adds that engaging in moral thinking is a tendency of the human mind that "I personally cannot help respecting deeply and I would not for my life ridicule it" (12).

Reflecting this, let us say that in prudential normativity the self-harm is *primary*—it is what *makes* the action imprudent.

Now let us contemplate the proposal that moral normativity might be identified with prudential normativity. (Note that I am not targeting the view that acting in morally bad ways is imprudent—Bloomfield's position—but rather the stronger identification claim that moral badness *is* imprudence.) Consider the Nazis, whose actions were so horrendous that even trotting them out endlessly as a philosophical example shouldn't dampen our horror at what they perpetrated. The error theorist may despise the Nazis as much as anyone, but nevertheless withholds assent from the claim that what they did was *morally* wrong. (Obviously, the error theorist needs to be careful in voicing this claim, for it is likely to be misconstrued as indicating some kind of tolerance for the Nazis, whereas in fact she simply thinks that all moral language is bankrupt: that the Nazis' actions were not morally wrong, not morally right, not morally permissible, not morally anything.) Let us focus our attention on a particular SS guard, who herded frightened Jewish children into the gas chambers in full knowledge of what he was doing. Let us stipulate that no possible defense could be mounted for his deeds; if any action is a moral crime, it is his.[14] Now let us adopt the proposal that the wrongness of his actions is nothing more than their imprudence (i.e., that moral badness is imprudence). This means accepting that what primarily makes the guard's action wrong is that he harmed *himself*. The fact that he harmed others contributes to the wrongness, but only derivatively (in that in harming others he harmed himself), and it is the harm to himself that really determines his wrongness. It also means accepting that what determines the *magnitude* of his crime is the magnitude of the injury he does himself (i.e., in harming so many innocent victims he damaged himself severely). Furthermore, it means accepting a counterfactual: that if the guard had killed all those innocent people *but had managed somehow (magically, perhaps) to avoid the consequential self-harm*, then there would have been nothing wrong (i.e., morally/prudentially wrong) with his actions.

This, I hope, sounds appalling. It might not be unreasonable for us to agree that the guard *did* harm himself in various ways, but the idea that the wrongness of his actions derives solely from that self-harm is a monstrous thought—almost as monstrous as the thoughts the guard uses to justify his actions to himself. The example illustrates the enormous difference between prudential and moral norms, and does so at an intuitive level, without pretending to articulate what a moral norm is. (We are supposed to think "Whatever exactly a moral norm is, it's not like *that*.") There are many ways to demonstrate the difference between these two types of normativity. To perform an action that harms oneself (e.g., to drink strong coffee before going to bed) may amount to doing something that one ought not to do, but it's not the right kind of "ought-not-ness" to count as a *transgression*—and the notion of transgressing is surely fundamental to moral thinking. The "emotional profiles" of prudence and morality appear intuitively to be very

---

14. The moral error theorist who thinks that moral predicates have empty extensions across all possible worlds will struggle to take this last conditional phrase literally as a counterfactual truth. I submit, however, that even she can understand the spirit of the claim, and treat it as an acceptable rhetorical pronouncement that stands in for some true complex proposition.

different. Our basic emotional response to someone's self harm is *pity*. The emotion of *retributive anger* makes little sense within the framework of prudential normativity, for what sense is there in the idea that someone who has harmed himself *deserves* the infliction of further harm (or, moreover, that the severity of the harm we inflict should be proportional to the degree of self-harm)?[15] Harming oneself *per se* doesn't (and shouldn't) provoke the emotion of *guilt*; it provokes the phenomenologically very different form of self-castigation of thinking "I'm so stupid" (and is *that* what we think the SS guard should be feeling?). Without underwriting guilt, it is implausible that prudential considerations could form the lifeblood of a moral *conscience* in the way that moral considerations do. Consider also the reparations that on many occasions we would insist that the moral criminal make to his victims. On the morality-qua-prudence view, the primary victim of any crime is always the criminal himself. Perhaps compensating the other victims (or simply apologizing to them) will be a means for the criminal to benefit himself, to undo the self-injury that he has inflicted, but there is no reason to assume that this is the only or the best way for him to accomplish this end, and thus if he finds some other way of compensating for the harm he did himself (taking a relaxing holiday? treating himself to a special gift? forgiving himself?) then this act of direct self-profit may well be the preferable course for him.

It may help to clarify my central claim—that moral badness and imprudence are non-identical—if it is observed that it is consistent with maintaining any or all of the following:

i. Performing actions of the types that are typically thought of as morally bad will cause self-harm.

ii. (i) is true always, necessarily, and for everyone.

iii. Moreover, the degree of self-harm is proportional to the magnitude of the (assumed) moral badness.

iv. Some actions are considered both morally bad *and* imprudent, making it sometimes difficult to tease the two apart.

v. When we try to dissuade someone from performing a morally bad action, the negative consequences that will befall him are likely to be among the first things we mention. (We may even have a deeply entrenched and institutionalized cultural tradition of appealing to the punishments of an all-powerful divine entity in order to back up our moral judgments, thus ensuring that we think of moral transgressions as imprudent.)

vi. Sometimes normative frameworks are "nested," such that one is obliged, according to framework A, to follow the prescriptions of framework B. Thus, in some circumstances a person may think it morally required to be prudent. This, again, makes it hard to tease the two apart, but does not indicate the absence of a

---

15. This is not to deny that there may be other grounds for punitive response for which the idea of *desert* plays no role. In punishing the SS guard we may hope to discourage him from harming himself in this manner again, or hope by example to discourage others from such heinous acts of self-harm.

distinction. (By analogy, a parental authority may decree to a child "Do what the teacher tells you to." If the teacher then orders "No talking," then we may say that not talking has been prescribed both directly by teacherly authority and indirectly by parental authority. But the two normative frameworks are nevertheless distinct, and their respective values and rules may have very different characteristics.)

vii. Moral norms need not be exclusively other-regarding.[16] The sentence "You ought not neglect your health" may be used to express a piece of prudential advice, or could be used to state a self-regarding moral imperative. These respective usages would display different characteristics. (If used morally, for example, the "ought" claim will make legitimate certain kinds of criticism for non-compliance that a prudential usage would not.)

viii. To observe the distinction between moral normativity and prudence is not to disparage prudence or suggest that it must take a back seat to morality.

The form of argument pursued above—examining the characteristics of a normative system that is being offered as a candidate for vindicating morality, and declaring that it displays insufficient mesh with our pretheoretical desiderata concerning what moral normativity is like—is a regular task for the moral error theorist; she will find herself doing it again and again. Defeating the candidacy of *prudence* is fairly undemanding, I think, and can be successfully accomplished while keeping the discussion at a rough, intuitive level. But the error theorist's task may not always be so easy, and for other claimants it may be necessary for her to spell out in as much detail as possible what she takes to be distinctive (and problematic) about morality, analyze carefully the characteristics of the candidate, and compare the two. The error theorist may accept that some candidates fare better than others—some may have a much better claim than prudence—but she believes that ultimately none comes close enough to deserve the name "morality." The closest satisfiable satisfier of all our moral desiderata still counts at best as "schmorality."

Let me be clear what is meant by "schmorality" in this context. Picture a continuum comprised of what can be thought of (in a benignly vague manner) as "normative frameworks." At one end we have value systems that clearly count as *moralities*: Christian ethics, deontological systems, Moorean intuitionism, Platonic theories about the Form of the *Good*, and so on. The error theorist doesn't doubt that these moralities exist, but she thinks that none of them deserves to be epistemically endorsed. At the other end we have things that clearly don't count as moralities: the rules of chess, etiquette, doing whatever the hell you feel like, and so on. The moral error theorist is free to epistemically endorse the claims of such systems (e.g., she thinks that "You must not move your knight in a straight line" is true). Somewhere on this continuum will lie normative frameworks for which it is not immediately apparent whether they count as moralities: Some people will think they do, others will think they don't. Call these items "contenders," of which one example is prudence. The error theorist, as we have seen, thinks that prudence is a

---

16. Cf. Falk (this volume).

poor contender for being a moral system. (Indeed, even those with no sympathies with moral skepticism should assent to this.) Note that calling prudence a poor contender for being a moral system is not to call it a poor moral system (which would imply that it is a moral system), any more than a hopeless contender for being elected president is thereby a hopeless president.[17] It is not that the error theorist fails to epistemically endorse prudence (she may agree that Ernie ought not eat cookies in bed); but rather she thinks— for the kinds of reasons outlined above—that there is simply insufficient mesh between prudential normativity and moral normativity for prudence to count literally as *a morality*. And the moral error theorist thinks this about *every* contender: either it may be epistemically endorsed but is too far from the "morality" end of the spectrum to count literally as a morality, or it is close enough to count as a morality but (for various reasons) cannot be epistemically endorsed. Indeed, holding this combination of views is constitutive of being a moral error theorist. Every contender is thought to be either unsuccessful—i.e., there is nothing in the world answering to its claims, there is nothing that renders these claims true—or a schmorality: something bearing a resemblance to a morality—enough, perhaps, to be mistaken for the real thing by the inattentive—but which falls short of really being so.

What determines whether something is a morality or a schmorality? In my opinion, the answer turns on how the concept *morality* is *used*. If concept A is used in a certain manner, but turns out to be problematic for various reasons (i.e., it is uninstantiated by the world), and concept B is an instantiated contender for replacing A, then B can be an adequate successor only if it too can be used in the same manner. For example, even when we realized that nothing is absolutely simultaneous with anything else, the relativistic notion of simultaneity was able to take over seamlessly, since it works just as well in everyday contexts for creatures whose movements don't approach a significant fraction of the speed of light. We can *use* the concept of relative simultaneity in the same way as we can use absolute simultaneity, which suggests that the change didn't amount to replacing one concept with a different concept at all, but rather we just made a revision internal to a single concept. Thus we are not forced to the radical position that every pre-Einsteinian assertion of two events occurring simultaneously is false. By comparison, when we discovered that there are no diabolical supernatural forces in the universe, we had no further use for the concept *witch*. Perhaps we could have carried on applying the word "witch" to women who play a certain kind of local cultural role on the margins of formal society—perhaps we might even have located a cluster of naturalistic properties that all and only these women have—but carrying on in this way would not have allowed us to *use* the word "witch" for the purposes to which we had previously put it: to condemn these women for their evil magical influence and justify their being killed. Thus, there was little point in persisting in using the word "witch" to stand for certain instantiated naturalistic properties; we dropped it and concluded that all historical

---

17. Someone once claimed to me, in all seriousness, that golf was his religion. The correct response is not that golf is a very poor religion, but that it doesn't count literally as a religion at all.

assertions that certain women were witches—even the loosely spoken ones—were false; we became error theorists about witches.[18]

The question, then, in the moral case, is "What do we use morality for?" The answer will almost certainly be extremely complex, and is, moreover, largely an empirical business. It is extraordinary how rarely this matter has been squarely faced, and deplorable that on those occasions that are an exception, vague intuitions from the armchair have, more often than not, been thought to suffice. And yet on this question, as we have seen, depends the issue of whether all our moral utterances are true or false. If a contender for satisfying our pretheoretical desiderata for *morality* turns out to be something that we couldn't even use for the purposes that we have customarily put moral discourse—if, for example, we couldn't use it to justify deserved punishment, if it couldn't undergird the emotion of guilt, if it couldn't act as a bulwark against a range of motivational infirmities—then we have good reason for thinking that we have in our hands but a *schmorality*. And if this is so of the *best* satisfiable candidate(s), then we should all be moral error theorists. Obviously, no deliberation of this kind can proceed until we know just what it is that we *do* use moral discourse and moral thinking for. Thus, until the jury delivers its verdict on this empirical matter, the fundamental metaethical disagreement between the moral error theorist and the moral success theorist (i.e., the cognitivist who believes that moral assertions are often true) remains at a stalemate.

Let me give one brief example of this kind of exchange, more for the sake of clarity than argumentative success. David Lewis offers a candidate for satisfying the noun "value": that "something of the appropriate category is a value if and only if we would be disposed, under ideal conditions, to value it" ([1989] 2000: 68). The interesting details need not detain us here; the important point is that one of the discomforting implications of Lewis's offering is that, since human psychology is contingent, we might have valued different things (even under ideal conditions), thus there could have existed values different from those that actually do exist. Lewis's gentle example is that we might have valued seasickness and petty sleaze, but obviously far nastier things could have turned out to be good, according to his theory. Lewis admits that this rampant relativism is a disturbing implication, yet still thinks that his offering may be "as near right as we can get" to satisfying our problematic moral notions, supporting the conclusion that although "strictly speaking" the moral error theorist wins the day, "loosely speaking" values exist (92-93). Lewis may be correct. But how can we tell? How do we know when "Close enough is good enough"? According to my thinking, we must ask whether Lewis's "values" can play the same practical roles in our lives as moral values hitherto have done. What is interesting about Lewis's discussion is that he himself suggests a use to which we put values—one that turns out to undermine the candidacy of his favored claimant. The telling moment comes when he suggests why it is that relativism "feels wrong": He says that perhaps it is because "a large and memorable part of our discussion of values consists of browbeating and being browbeaten.[19]…The rhetoric would fall flat if we

---

18. This paragraph is lifted from my 2006, chapter 6.
19. Lewis here footnotes Hinckfuss (1987).

kept in mind, all the while, that it is contingent how we are disposed to value" (92). Lewis's intention is to diagnose the source of our uneasiness about relativism, but if we take seriously the thought that such rhetorical impact is an important part of the use to which we put moral considerations (both interpersonally and, perhaps, intrapersonally), then he has provided us with evidence *against* the adequacy of his theory of value, since he has identified an important practical purpose that would be lost if we adopted his replacement concept. (It is perhaps a depressing thought that this might be a central function of moral discourse, but, as I declared above, this is something for which hopeful or romantic guesses won't stand in for evidence.[20]) Thus there is at least one consideration—by Lewis's own lights—in favor of thinking that his "values" are not the real McCoy, in favor of thinking that he has provided us with a schmorality rather than a morality.

Suppose the error theorist is correct in holding that the closest satisfiable claimants for our moral concepts are all schmoral concepts. The question arises as to what she then does with moral concepts. The natural assumption is that the error theorist will also be an eliminativist: that she will recommend the abolition of moral language in all unembedded positive contexts. (These last qualifications are supposed to indicate that nobody thinks that we should eliminate moral language *altogether*; the error theorist will still assert things like "There exists nothing that is morally bad" and "St. Augustine believed that stealing pears was morally wrong.") The popular assumption is that if we catch a professed moral error theorist employing moral talk then we can triumphantly cry "Aha!" and accuse her of committing the intellectual vice that Quine (in a tone of disgust) characterized as engaging in "philosophical double talk which would repudiate an ontology while simultaneously enjoying its benefits" (1960: 242). Any such accusation is an argument not against the moral error theory but against the theorist—showing her to be a hypocrite, disingenuous, in bad faith, or vacillating between belief and disbelief. (Perhaps, on the latter charge, the error theorist is like Hume's Pyrrhonian, who, it will be recalled, cannot live his skepticism because "nature [is] too strong for it" ([1740] 1978: 657).)

But eliminativism does not follow logically from the error theory. The question of what one *ought to do* with one's moral discourse need not be a moral inquiry but may be construed as a practical question: Perhaps it involves a prudential "ought," or perhaps a hypothetical "ought" concerning how the agent's (idealized and fully-informed?) desires may be optimally satisfied.[21] I don't intend to adjudicate on this matter; all that is of concern here is that it is a kind of practical question that (we have allowed) the moral error theorist has the resources to address. Let us just say that the error theorist will opt to

---

20. Of the uses to which we put morality, to ignore some, in this calculation, on the grounds that they are considered "immoral" would, obviously, be to beg the question against the moral error theorist.

21. These disjuncts are distinct on the assumption that psychological egoism is false (an assumption that I feel confident in making). The falsity of psychological egoism means that a person—even a moral skeptic—may have genuinely non-derivative desires for others' welfare. Any "ought" claim that constitutes advice on how such an altruistic desire is best satisfied need not correspond to a prudential "ought."

eliminate moral discourse only if that conclusion is supported by some kind of cost-benefit analysis in comparison with other options. Yet what are the other possible options figuring in this calculation? The option of carrying on as if nothing has changed—of continuing to assert moral propositions and to hold moral beliefs *even while maintaining moral error theoretic commitments*—is surely a non-starter, for the kind of doxastic schizophrenia involved in such a life not only violates epistemic norms, but can be expected to lead to various kinds of pragmatic handicap. But there is a third option: The error theorist may consider taking a fictionalist attitude towards morality. The fictionalist's point of departure is summed up nicely by Hans Vaihinger:

> An ideal whose theoretical untruth or incorrectness, and therefore its falsity, is admitted, is
> not for that reason practically valueless and useless; for such an idea, in spite of its
> theoretical nullity, may have great practical importance. (1935: viii)

To adopt a fictionalist stance towards morality is to continue to make moral *utterances* and have moral *thoughts*, but withhold assertoric force from the utterances and withhold doxastic assent from the thoughts. The fictionalist can be seen as an error theorist who attends to both epistemic and pragmatic norms.[22] His respect for epistemic norms means that he steadfastly refuses to *believe* any moral claim; his sensitivity to pragmatic norms means that he seeks and embraces the expedient option.[23] On the assumption that morality is in various respects useful when it is asserted and believed, eliminativism will (ceteris paribus) constitute a practical cost; and if morality is *very* useful then eliminativism will constitute a *big* cost. The fictionalist option, therefore, becomes attractive if (and only if) it promises to recoup some of these costs. The advocate of fictionalism holds that some of these losses may be recovered by adopting a policy of employing moral language, engaging in moral deliberation, and being moved by moral emotions, but throughout it all remaining disposed to deny the truth of any moral proposition if pressed in an appropriately serious manner (e.g., when in the philosophy classroom), thus not really believing any of it (thus not violating any epistemic norms), and thus deflating a host of well-thumbed philosophical problems concerning the ontology of moral facts and our access to them. Regarding actual moral discourse the fictionalist remains an error theorist: He thinks that this discourse does aim at the truth but systematically fails to secure it. On the grounds of expediency he advocates a

---

22. Sometimes the label "fictionalist" refers to a philosopher advocating that we adopt a fictive stance; sometimes it refers to someone who has adopted that stance. (If certain critics of fictionalism are correct, there are no fictionalists in the latter sense.) Though potentially confusing, this equivocation seems benign in most contexts.

23. I should like to draw attention again to the distinction observed in note 21. An error theorist may have reason to adopt the fictive attitude because doing so promises to satisfy certain of her (idealized and fully-informed?) desires—and I see no grounds for denying that (some of) these desires may be genuinely altruistic in content. Thus in fact it need not be self-interest that recommends the adoption of the fictive stance. Nevertheless, counsel that appeals to self-interest is more likely to have a broader general influence, and thus (giving consideration also to the demands of concision) I will continue to fudge over this subtlety, and speak as if self-interest were the only relevant consideration motivating the fictionalist.

revolution in our attitudes towards morality, and regarding the (imaginary) post-revolution moral discourse the fictionalist is no error theorist, for, come the revolution, moral discourse will no longer aim at the truth.[24] The tricky part of expounding fictionalism is to make out a kind of attitudinal acceptance *other than belief* that can play a central role in serious intellectual inquiry and serious practical deliberation.

There are many objections to fictionalism in general, and some to moral fictionalism in particular. (For discussion, see Hussain 2004; Kalderon 2005*a*, 2005*b*; Nolan et al. 2005.) In what remains I will discuss three objections that are similar in that each holds that moral fictionalism somehow undermines the error theory on whose shoulders it stands, thus rendering itself redundant (in the sense that if the error theoretic account of Xs becomes implausible, then although taking a fictive attitude towards Xs remains an intelligible option, there is no need to do so). The first two objections can be interpreted as maintaining that anyone attempting to fictively accept morality must be epistemically endorsing a morality after all. The third objection doesn't quite amount to this, but is related in that it holds that embracing (a particular kind of) fictionalism will destabilize a particular kind of argument in favor of the moral error theory.

First, one might complain that if the policy of uttering and thinking moral propositions is recommended on prudential grounds, then moral discourse has been vindicated after all. Indeed (the complaint might continue), the fictionalist has supplied evidence against his own error theory, since he has provided grounds for equating moral norms with prudential norms. This is somewhat different from the objection to moral error theories that we encountered earlier. Then the claim was that if the error theorist agrees that *acting* in accordance with assumed moral norms is justified on prudential grounds, then he has provided morality with all the justification that it needs. Now the claim is that if the error theorist agrees that *talking and thinking* in moral terms is justified on prudential grounds, then he has provided morality with all the justification that it needs.[25] But the response is much the same. We should start by bearing in mind the distinction between epistemic justification and instrumental justification. If someone holds a gun to your head and says "Utter the sentence '1 + 1 = 3' or I'll shoot!" then the act of utterance will be prudentially wise (instrumentally justified), but the content of the utterance will be no less false—and any act of believing it no less illegitimate (no more epistemically justified)—for that. Recall that the error theorist has been impressed by the thought that moral propositions have substantive metaphysical (and problematic) implications or presuppositions which prudential propositions lack. The fact that the act of uttering one of these flawed sentences may be instrumentally justified hardly shows that the sentence must be true, or that believing the sentence would be epistemically justified. Nor does the fact that uttering a normative sentence is prudentially justified

---

24. The kind of fictionalism being described here is the "revolutionary" branch. In contrast, a "hermeneutic" fictionalist argues that we have been taking a fictive attitude towards the target discourse all along (and thus the hermeneutic fictionalist is not an error theorist). Hermeneutic moral fictionalism is advocated by Mark Kalderon (2005a) and criticized by Jason Stanley (2001).

25. Crispin Wright may be interpreted as presenting an argument along these lines. See his 1992: 10; 1996: 3.

mean that the sentence really expresses nothing other than a prudential norm. On this last point it might be useful to consider a comparison. A person might choose to cultivate the personality trait (assuming that it deserves to be so-called) of having altruistic emotions towards his friends and family.[26] Quite how one goes about such an act of "cultivation" need not bother us now; the point is that some act of deliberate choice is involved, which, if successful, results (at some time in the future) in having interests in the welfare of certain others—interests that do not depend on the contribution that the others' welfare makes to one's own interests. The important thing to notice is that at the time of original deliberation the person may be calculating entirely in selfish terms; she may realize that having altruistic emotions will, in various ways, contribute to her own welfare. This observation, however, in no way undermines the possibility that the love and sympathy that this person eventually comes to feel are genuinely altruistic in nature. One can be selfishly motivated to become a less selfish person, and may succeed. Similarly, one can be motivated on grounds of self-interest to adopt a policy of accepting a certain class of normative claims—which are distinct from prudential claims—and may succeed.[27]

The objection just discussed was that moral fictionalism undercuts its own error theoretic basis—that adopting a fictive attitude towards morality amounts to an epistemic endorsement of it—and thus if one wants to maintain a moral error theory one had better eschew fictionalism, which more or less amounts to advocating that the error theorist be an eliminativist. The second objection is that the fictionalist stance is incoherent because the distinction between belief and "acceptance" cannot be maintained (see Putnam 1971: 68-69; Newman 1981). On this view, if someone acts, talks, thinks and feels in accordance with having moral beliefs, then he actually *does* have moral beliefs. Thus, this objection also amounts to the allegation that attempting to adopt a fictive attitude (about anything this time, not just morality) will amount to an epistemic endorsement, and that if one wants to be an error theorist one had thus better steer clear of fictionalism.

Since *belief* is a contested notion, the suspicion arises that some accounts of belief will allow for a separate category of *acceptance* while others—e.g., neo-behaviorism— will not. And so it may seem that the only means of responding to this objection is to provide a convincing argument for one of the former accounts. But in fact there is good reason for thinking that *all* parties have cause to allow this distinction, even the neo-behaviorist. Consider the crudest kind of behaviorism that says that all it is to believe that *p* is to act as if one takes *p* to be the case. Even so boorishly extreme a behaviorism will want to allow that on occasions a person may act as if she takes *p* to be the case without believing that *p*. Actors, for example. This observation alone forces the acknowledgement of some category of acceptance distinct from belief: It is the attitude actors take towards elements of the fiction into which they enter. The thing about actors, of course, is that they are disposed to "step out" of the fiction; they don't act *all the time* as if they take *p*

---

26. Note that here I am considering altruistic emotions in a non-moral sense. To *like* someone—to have a non-derivative concern for his welfare, to be motivated to act to further his interests, to feel affection towards him—is a capacity that might be enjoyed by a creature entirely lacking the cognitive sophistication to make any moral judgments at all. (See notes 9 and 10.)
27. Compare David Schmidtz (1994) on "maieutic ends."

to be the case. But the crucial detail to notice about the fictionalist is that he too remains disposed to step out. There are contexts where he does *not* speak as if he takes *p* to be the case: namely, when he is in the critical context of declaring his endorsement of the moral error theory.

But the person pressing this objection may persist. Even though acknowledging some kind of attitude—distinct from belief—deserving the name "acceptance," she may doubt that one can be in this state with respect to some subject matter *nearly all the time*. Our crude behaviorist may revise slightly: All it is to believe that *p* is to act *at least nearly all the time* as if one takes *p* to be the case. The idea that the matter might depend on the *amount of time* one spends "immersed" in the fiction compared to the amount of time one spends "outside" it, strikes me as terribly improbable. What constitutes "nearly" here? Even acknowledging that the answer may be vague ("Around 90%"), it seems crazy to think that if I spend 95% of my time acting as if *p* were the case then I believe that *p* but if I spend only 85% of my time acting in this way—all else remaining the same—then I do not believe that *p*. (I'm sure there's a better objection to this than "Horrible theory!" but I'm happy on this occasion to rest matters there and trust in the reader's agreement.)

Far preferable would be the provision of some account of the *nature* of the two kinds of context, such that we can see that in one context utterances match what one really believes, even if it is a context entered into very rarely in comparison with the other context. Above I called the context of expressing disbelief (e.g., when doing metaethics) the more "critical" context, and this is the term I have used on other occasions (Joyce 2001, 2005). It is, perhaps, an ill-chosen word, since it suggests that there is something "uncritical" about the fictionalist's engagement with moral matters in everyday life. It is important to see that "critical" here is a term of art, indicating an asymmetrical relationship between the two kinds of context (or, rather, naming a pole at one end of a continuum of contexts). Context *n* is more critical than context *m* if and only if *n* is characterized by a tendency to scrutinize and challenge the presuppositions of *m*, but not vice versa. This is consistent with *m* being the more "critical" in a vernacular sense of the word. For example, working out the plot of a complex novel may involve a great deal of careful thinking, whereas the thought "It's all just a fiction" is a simple matter. Nevertheless, in the sense intended, the latter is the more "critical" context since it questions the world of the novel. Similarly, when immersed in morality the fictionalizing error theorist may deliberate extremely carefully about consequences, weigh outcomes thoroughly, deploy acute powers of imagination and reflection, and so on, and yet still not inhabit his "most critical context" where he denies moral truth across the board. Though this amounts to not much more than a gesture—most prominently leaving us wondering just what is meant by "scrutinize and challenge"—I believe it is a promising way of addressing the problem, which, if successful, will make the *amount of time* one spends in the critical context irrelevant to the question of what one believes.

Those who doubt the viability of the belief/acceptance distinction may have their skepticism alleviated if they reflect on the seeming ubiquity of the phenomenon—or, at least, of closely related phenomena. The human proclivity for engaging with fiction

(novels, movies, etc.) is the most conspicuous example, but arguably there are many less obvious instances of similar mechanisms operating in everyday life. Michael Bratman (1992) has argued persuasively that all practical reasoning involves accepting (but not believing) certain propositions as a background to effective deliberation. On the assumption that psychological simulation involves a kind of acceptance-without-belief, acceptance may be implicated in hypothetico-deductive reasoning, ascribing mental states to others, and predicting others' behavior (see discussion in Davies and Stone 1995*a*, 1995*b*). Simulation probably plays a central role in empathy (Goldman 1992) and visual imagination (Currie 1995). Vaihinger (1935) supplies numerous mundane examples of our treating something "as if" it were true while knowing that it is not. (While Vaihinger almost certainly errs on the side of over-enthusiasm, his catalog of examples of the fictional stance is nonetheless instructive.)[28]

The fictionalist's strategy here is unashamedly one of finding partners in innocence. Although it is unlikely that there is a single belief/acceptance distinction that all the aforementioned phenomena exhibit, there is enough family resemblance here that it is not unreasonable for the fictionalist to think that by cozying up *his* kind of belief/acceptance distinction to these other commonplace examples he can dispel some knee-jerk doubts. Is someone who reads a novel disingenuous or self-deceived? Is someone who engages in role-play suffering from anything deserving the name "schizophrenia"? Is someone who accepts a proposition as a background assumption when deliberating manifesting bad faith? Does feeling empathy make one a hypocrite? And does engaging any or all of these practices have deleterious effects on one's interests? I take it that the answer to all is "No." Now, as admitted, the kind of belief/acceptance distinction at the heart of the moral fictionalist's case may not be quite the same as these other instances, but its similarity to these "innocent" examples is sufficient at least to show that such accusations (that the practicing fictionalist is in bad faith, suffers from self-deception, etc.) cannot be pronounced lightly. The onus, of course, is first on the fictionalist to articulate with precision what the distinction he has in mind amounts to; accusations of bad faith, schizophrenia, etc. must be suspended until then. Then, of course, the burden falls to the opponent of fictionalism to replace the vague rhetorical sense of terms like "bad faith" and "schizophrenia" with something literal (and obviously undesirable).

The third objection to fictionalism that I will briefly comment on targets a particular brand of moral fictionalism—but since it is a kind that I have on occasion defended (Joyce 2001, 2005), I feel moved to respond. The fictionalism in question is one that hypothesizes that engaging in moral discourse is useful in a particular way: namely, that this engagement stimulates *motivation* in a pragmatically desirable manner. (Any fictionalist theory that assigns a different sort of usefulness to morality will not be affected by this objection.) It may be hypothesized, for example, that the expediency of

---

28. There are other "belief versus acceptance" distinctions in the philosophical literature that probably have little to do with the phenomenon (or family of phenomena) that is relevant here. Bayesian decision theorists often distinguish between partial belief and full acceptance (see Swain 1970). There is debate about whether collectives of individuals can have belief or merely acceptance (see Wray 2001). See also Cohen 1992; Frankish 1998.

moral discourse derives from its capacity to act as a bulwark against various kinds of practical infirmity—e.g., weakness of will, discounting future gains, and so on—better than clearheaded instrumental deliberation. Thinking of an action as something that "just *must* be done" may encourage performance of that action more reliably than explicitly conceiving of the action as one that serves one's long-term best interests; imagining the omission of that action to be something that will not merely frustrate one's desires but make one *reprehensible* and *deserving of punishment* may be more likely to result in resolve to perform the action. This, it seems to me, is an intuitively attractive idea, especially when it is made clear that the moral judgment may come "embedded" in an emotion, such as guilt or punitive anger. There is plenty of empirical evidence that self-directed moral emotions have motivational efficacy (see Carlsmith and Gross 1969; Freedman 1970; Tangney and Dearing 2002; Ketelaar and Au 2003; Zhong and Liljenquist 2006; Tangney et al. 2007).

So what problem does this hypothesis pose for fictionalism? There is of course the burning question of how taking a *fictive* attitude towards a set of norms and values could possibly engage motivation in this way. But that is an empirical question that I don't propose to discuss here (see Joyce 2001, 2005, 2006); rather, I am interested in the theoretical question of whether supporting this hypothesis makes trouble for the error theoretic basis of fictionalism. There are two reasons for thinking that it might. The first is that if a moral judgment engages motivation in this manner, then doubt is cast on the claim that moral judgments are a cognitive affair. But if moral judgments are in fact a *noncognitive* affair, then the moral error theory collapses, for one of the distinguishing features of this metaethical theory is its commitment to cognitivism.

This objection is confused. Noncognitivism is a thesis about what kind of mental state(s) moral judgments express; it denies that the state expressed is *belief* (i.e., it denies that moral judgments are assertions). One popular form of noncognitivism—emotivism—claims that what is expressed is some (specifiable) conative or emotional state. To advocate cognitivism, however, is not to make the wild claim that moral judgments have nothing to do with emotions. Cognitivism is compatible with the view that moral judgments reliably prompt emotional activity. It is compatible with the view that moral judgments generally, or even always, flow from seething emotional activity in the brain. It is compatible with the view that what goes on when one makes a moral judgment is that one "projects" one's emotional life onto the events of the world. It is compatible with the view that the human capacity for moral judgment is a discrete biological adaptation which evolved precisely by virtue of its tendency to affect human emotions in a fitness-enhancing manner. None of these possibilities—nor, indeed, all of them jointly—entail the denial of the claim that moral judgments are assertions. Cognitivism is compatible even with the claim that the connection between moral judgments and emotional activity is a *necessary* one (though I should add that this is not a claim I endorse). Consider, by analogy, the act of promising. The criteria for a promise to have occurred involve a range of linguistic conventions in which both promiser and promisee need be versed; for example, the addressee must hear and understand the words uttered, and the speaker must take it that this is the case. If the addressee doesn't hear the "I promise…" claim, or

doesn't understand what the utterance signifies, then the act of promising misfires, and no promise has occurred (see Austin 1962). The satisfaction of these criteria will require both speaker and addressee to have certain *beliefs*—for example, the speaker must believe that his addressee hears and understands. This connection is a necessary one: It is not possible that any person could succeed in making a promise to another person without having such a belief. And yet we would hardly say that the act of promising *functions to express the belief that one's audience hears and understands* (rather, a promise expresses an intention). Therefore, since a kind of speech act and a mental state may be necessarily linked without the former *functioning to express* the latter, then even if it were the case that moral judgments necessarily engage motivational states, noncognitivism is not the automatic conclusion (see Joyce 2002).

The second potentially problematic implication of assuming that the usefulness of moral thinking lies in its impact on motivation is even more limited in scope: It is a problem only for the error theorist who has employed a particular kind of argument to establish her error theory. Several philosophers who harbor skeptical misgivings about morality derive their doubts (in part) from a commitment to a Humean psychology, according to which beliefs and desires are distinct and but contingently linked states (see Williams 1981; Mackie 1977; Joyce 2001). John Mackie, for example, thinks that moral imperatives imply external reasons claims[29] (to import Bernard Williams's terminology), but, like Williams, he thinks that all legitimate reasons claims are internal. (An internal reason is one that suitably connects with a person's "subjective motivational set"; an external reason is one that does not.) The basis of this latter opinion (for Williams at least) is the thought that any reason must potentially motivate a person, but only internal reasons can do so—and his ground for thinking this is that believing oneself to have an external reason cannot (alone) prompt motivation, because no belief can do that.

It should be stressed again that one might be a moral error theorist on grounds having nothing to do with any of these considerations. But if one *is* moved by this argumentative thread to adopt a moral error theory, and if, in addition, one is moved to become a fictionalist by the thought that morality is useful *because of its influence on motivation*, then, it may be supposed, one has some explaining to do. (It may sound like a fairly specialized position that is being defended—and indeed it is so—but it is not an uninhabited position, and the objection has been raised on more than one occasion.[30]) How could a moral *belief* (understood in Humean terms) function to influence motivation? In fact the explaining is fairly easy. The hypothesis that moral judgments are useful because they influence motivation need involve only the claim that moral judgments *often* or *reliably* or *defeasibly* or *contingently* engage motivational structures. (Indeed, even "sometimes" will suffice.) It is perfectly possible that moral beliefs are just that—beliefs—and that beliefs alone never prompt motivation; but it may also be claimed that such beliefs, when they figure in an ordinary person's psychological economy—an

29. See Mackie 1982: 115.
30. Yes, I'm afraid I've been reading the reviews of my own book (2001): see McKeever 2003 and Wallace 2003. See also Hussain 2004.

economy that includes typical desires and emotions—will generally have an impact on motivation.

Having deflected these criticisms (and finding no other compelling), I feel confident in claiming that the fictionalist position is at least *coherent*. Whether the fictionalist stance is psychologically feasible, and whether it will supply the promised pragmatic gains, remain serious empirical uncertainties. Though on other occasions I have advocated the case for fictionalism (Joyce 2001, 2005), it must be underscored—as I did at the time—that while there is certainly a place for plausible speculation when it comes to directing people's attention, nothing confident can be claimed in advance of the *a posteriori* footwork. Perhaps in the end the data will not favor the fictionalist option. Perhaps eliminativism will be the better course for the moral error theorist, in which case she may use the term "schmorality" in its customary pejorative sense: to scoff "*Morality, schmorality!*" But it is important to remind ourselves that even the eliminativist error theorist will still have plenty of good and strong reasons—many of them self-interested reasons—for being nice to her fellows.

REFERENCES

Austin, J.L. 1962. *How to Do Things with Words*. Oxford: Oxford University Press.

Blackburn, S. 1984. *Spreading the Word*. Oxford: Clarendon Press.

Bratman, M.E. 1992. "Practical reasoning and acceptance in a context." *Mind* 101: 1-15.

Burgess, J.P. 1978. "Against ethics." Unpublished manuscript. Department of Philosophy, Princeton University.

Carlsmith, J. and A. Gross. 1969. "Some effects of guilt on compliance." *Journal of Personality and Social Psychology* 11: 232-39.

Cohen, L.J. 1992. *An Essay on Belief and Acceptance*. Oxford: Clarendon Press.

Currie, G. 1995. "Visual imagery as the simulation of vision." *Mind and Language* 10: 25-44.

Davies, M. & Stone, T. (eds.). 1995*a*. *Folk Psychology: The Theory of Mind Debate*. Oxford: Blackwell.

Davies, M. & Stone, T. (eds.). 1995*b*. *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.

Frankish, K. 1998. "A matter of opinion." *Philosophical Psychology* 11: 423-442.

Freedman, J. 1970. "Transgression, compliance, and guilt." In J. Macaulay and L. Berkowitz (eds.), *Altruism and Helping Behavior*. Academic Press.

Garner, R.T. 1994. *Beyond Morality*. Philadelphia: Temple University Press.

Goldman, A. 1992. "Empathy, mind, and morals." *Proceedings and Addresses of the American Philosophical Association* 66: 17-43. Reprinted in Davies & Stone 1995*b*.

Hägerström, A. 1953. *Inquiries into the Nature of Law and Morals*. Stockholm: Almqvist & Wiksell.

Haji, I. 1998. *Moral Appraisability*. Oxford: Oxford University Press.

Haji, I. 2003. *Deontic Morality and Control.* Cambridge: Cambridge University Press.

Hinckfuss, I. 1987. "The moral society: Its structure and effects." *Discussion Papers in Environmental Philosophy* 16. Canberra: Philosophy Program (RSSS), Australian National University.

Hume, D. [1740] 1978. *A Treatise of Human Nature*. L.A. Selby-Bigge (ed.), Oxford: Clarendon Press.

Hussain, N.J.Z. 2004. "The return of moral fictionalism." *Philosophical Perspectives* 18: 149-187.

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

Joyce, R. 2002. "Expressivism and motivation internalism." *Analysis* 62: 336-344.

Joyce, R. 2005. "Moral fictionalism." In Kalderon 2005*b*: 287-313.

Joyce, R. 2006. *The Evolution of Morality*. Cambridge, MA.: MIT Press.

Kalderon, M.E. 2005*a*. *Moral Fictionalism*. Oxford: Oxford University Press.

Kalderon, M.E. (ed.). 2005*b*. *Fictionalism in Metaphysics*. Oxford: Oxford University Press.

Ketelaar, T. & Au, W.T. 2003. "The effects of feelings of guilt on the behavior of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction. *Cognition and Emotion* 17: 429-453.

Lewis, D.K. [1989] 2000. "Dispositional theories of value." In his *Papers in Ethics and Social Philosophy*. Cambridge: Cambridge University Press.

Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin Books.

Mackie, J.L. 1982. *The Miracle of Theism*. Oxford: Clarendon Press.

McKeever, S. 2003. "Review of Joyce, *The Myth of Morality*." *Ethics* 114: 182-184.

Newman, J. 1981. "The fictionalist analysis of some moral concepts." *Metaphilosophy* 12: 47-56.

Nolan, D., Restall, G. & West, C. 2005. "Moral fictionalism versus the rest." *Australasian Journal of Philosophy* 83: 307-330.

Putnam, H. 1971. *Philosophy of Logic*. New York: Harper & Row.

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

Schmidtz, D. 1994. "Choosing ends." *Ethics* 104: 226-251.

Stanley, J. 2001. "Hermeneutic fictionalism." *Midwest Studies in Philosophy* 25: 36-71.

Swain, M. (ed.). 1970. *Induction, Acceptance, and Rational Belief*. Dordrecht, Holland: D. Reidel.

Tangney, J., & Dearing, R. 2002. *Shame and Guilt.* New York and London: Guilford.

Tangney, J., Mashek, D., Stuewig, J., Magaletta, P. et al. 2007. "Working at the social-clinical-community-criminology interface: The GMU inmate study." *Journal of Social and Clinical Psychology* 26: 1-28.

Vaihinger, H. 1935. *The Philosophy of 'As If.'* C.K. Ogden (trans.), London: Routledge & Kegan Paul Ltd.

Wallace, R.J. 2003. "Review of Joyce, *The Myth of Morality*." *Notre Dame Philosophical Reviews* 2003.11.04.

Williams, B. 1981. "Internal and external reasons." In his *Moral Luck*. Cambridge: Cambridge University Press.

Wittgenstein, L. 1965. "Lecture on ethics." *Philosophical Review* 74: 3-12.

Wray, K.B. 2001. "Collective belief and acceptance." *Synthese* 129: 319-333.

Wright, C. 1992. *Truth and Objectivity*. Cambridge, MA.: Cambridge University Press.

Wright, C. 1996. "Truth in ethics." In B. Hooker (ed.), *Truth in Ethics*. Oxford: Blackwell.

Zhong, C. & K. Liljenquist. 2006. "Washing away your sins: Threatened morality and physical cleansing." *Science* 313: 1451-52.