

THE ERROR IN “THE ERROR IN THE ERROR THEORY”

Richard Joyce

[submitted 09/09 – comments still welcome]

The moral error theorist thinks that moral language and moral thinking aim at the truth (i.e., that moral language is assertoric and moral thinking doxastic) but that they systematically fail to secure the truth. This radical form of moral skepticism takes an attitude towards morality like sensible people take towards astrology, like reasonable people take towards talk of witches, and like atheists take towards religion. The moral error theory clashes not just with common sense, but with an unusually precious chunk of common sense, the undermining of which many people contemplate with deep unease. It is no surprise, then, that the error theorist faces opposition at every turn. Stephen Finlay [2008] joins the ranks of opposition with his recent paper “The error in the error theory”—not only inviting a rejoinder from the error theorist, but more-or-less forcing the skeptic’s hand in choosing a title for that rejoinder.¹ Having discharged the latter duty (see title, above), let me take on the former task. My aim in this paper is not merely to rebut Finlay’s critique, but to clarify what it is that the moral error theorist is committed to and how he or she might argue for it.

Finlay states that the moral error theory “is premised on two claims” [2008: 347²]:

- 1) *Presupposition*: moral judgments involve a particular kind of presupposition which is essential to their status as moral;
- 2) *Error*: this presupposition is irreconcilable with the way things are.

Allowing *Error*, Finlay sets himself the task of attacking *Presupposition*. (In this he is not alone; see Dreier, Kirchin, and West—all in Joyce and Kirchin (eds.) [2009]—Prinz [2007], Lewis [1989/2000], Shafer-Landau [2005], and many others besides.) This is what I have called “a concessive strategy” against the moral error theory: conceding that the component of morality that the error theorist maintains is hopelessly flawed is indeed hopelessly flawed, but nevertheless aiming to avoid the skeptical conclusion by insisting that moral discourse was never properly committed to this problematic component in the first place. (By comparison, the strategy of accepting (1) and denying (2) I call the “head-on” strategy.)

Finlay’s opening statement that the moral error theory “is premised on” (1) and (2) is ambiguous. If he means that (1) and (2) are premises for a well-known argument for the moral error theory then he is correct. If, however, he means that being an error theorist requires the endorsement (1) and (2) then he is incorrect. (1) and (2) together constitute *an argument for* the moral error theory. Perhaps it is an unsound argument, but there may be other sound arguments for that position. The failure of (1) or (2) does not imply the falsity of the moral error theory—any more than the failure of the Argument from Evil would show that God exists.

So the overarching error in Finlay’s paper is his confusing of the moral error theory with the argument(s) one might use to support it. He focuses on one positive argument for the moral error theory: a “pull factor” (as it were). But we should also be aware that one can be “pushed” to adopt a moral error theory through a dissatisfaction with all alternatives, and this

¹ Alas, Jonas Olson beat me to the punch, with his title to Section 4 of Olson 2010.

² Page references otherwise unqualified are all to Finlay 2008.

is a motivation that Finlay ignores. Suppose one first becomes convinced of the implausibility of noncognitivism: one thinks that moral judgments are truth-apt. Therefore one seeks an understanding of the requisite truth-makers. But suppose that when one examines all attempts to explicate moral properties, both objectivist and non-objectivist—when one critically scrutinizes intuitionism, constructivism, rationalism, various species of naturalistic realism, response dependency theories, and all the rest (including Finlay’s own preferred theory)—one finds serious problems and weaknesses with every contender. At a certain point it becomes reasonable to exclaim “Enough already!” One may at this juncture reconsider the prospects of noncognitivism, but assuming that the paucity of its attractions remains obvious, it would now be reasonable for one to become an error theorist.

Notice that no version of Finlay’s positive (1)+(2) argument would have played any part in this process; the process is one of becoming disillusioned with the multifarious attempts to epistemically vindicate morality, and eventually becoming seriously pessimistic about that whole program. It will be objected that if this is the process by which one becomes an error theorist, one could never be sure that a successful contender won’t show up next week, and thus one’s moral skepticism will be provisional. Quite so. One’s acceptance of the error theory will be proportional to one’s confidence that a successful vindication is not going to arrive. What I am suggesting is that after flicking through a back catalog of failed theories that extends back over two thousand years and involves some of the subtlest minds that humanity has to offer, one might have a fairly high confidence of this.

Not only does Finlay overlook how one might be pushed into skepticism, he is critical of only one kind of positive argument. It is widely reported that John Mackie offers two positive arguments for his error theoretic view: the Argument from Relativity and the Argument from Queerness. In fact, Mackie explicitly mentions five considerations:

[F]irst, the relativity or variability of some important starting points of moral thinking and their apparent dependence on actual ways of life; secondly, the metaphysical peculiarity of the supposed objective values, in that they would have to be intrinsically action-guiding and motivating; thirdly, the problem of how such values could be consequential or supervenient upon natural features; fourthly, the corresponding epistemological difficulty of accounting for our knowledge of value entities or features and of their links with the features on which they would be consequential; fifthly, the possibility of explaining, in terms of several different patterns of objectification ... how even if there were no such objective values people not only might have come to suppose that there are but also might firmly persist in that belief. [Mackie 1977: 49]

The relationship of Finlay’s (1)+(2) argument to these five is complicated, but one can say with confidence that the failure of *Presupposition* would have no effect on Mackie’s first argument—the Argument from Relativity (a.k.a. the Argument from Disagreement)—and probably no effect on the rarely-discussed last argument—the argument from objectification (projectivism)—either.³

It is tempting to see Finlay’s (1)+(2) argument as the form of the Argument from Queerness—but that is not quite right. Finlay’s *Error* is consistent with the presupposition in

³ I discuss the contribution of the argument from objectification to Mackie’s moral error theory in Joyce [2009]. The objectification thesis may well include a claim about the objective quality of moral *phenomenology*, but this appears to be logically independent of Finlay’s *Presupposition* thesis, which, on the face of it, is a claim about moral concepts. Moral phenomenology might have an objective quality without objectivity being written into the meaning of moral terms; and, conversely, moral concepts might include a claim of objectivity without this objectivity figuring in the way moral matters *seem* to us phenomenologically.

question being rejected simply on the grounds of empirical inadequacy. It is entirely conceivable that a moral error theory might be based on such an argument. For example, it might be that moral discourse presupposes a view of human autonomy which we discover *through empirical methods* to be flawed [see, e.g., Libet 2004]; or perhaps it presupposes a view of human character traits that we discover *through experimental psychology* to be wrong [see Doris 2002]; or perhaps it presupposes the truth of a kind of internalism that *neuroscience and psychopathology* reveal to be mistaken [see Roskies 2003]. Such arguments would satisfy Finlay's (1)+(2) format, but arguably none would count as an Argument from Queerness. In an Argument from Queerness, the presupposition in question is held to fail because it is too bizarre, too ontologically profligate, too far-fetched—just too darn queer—to be taken seriously. There are two things to note about this observation. First, it's rough; it employs indeterminate and relative terms like “bizarre” and “far-fetched.” Second, it's generic: It is a *type* of argument; the exact nature of the queer presupposition is left open, and there may be different ways of filling it in. We can speak of “*an* Argument from Queerness” of which one instance is “*Mackie's* Argument from Queerness.”

This, then, is yet another limit that must be noted about Finlay's case. All his efforts are spent on *Mackie's* Argument from Queerness (and, to the extent that it is distinct, *Joyce's* Argument from Queerness)—an argument for which the presupposition is (Finlay thinks) “absolute authority.” It is not that this is unreasonable—Mackie's Argument from Queerness is a central piece of skeptic weaponry (and I would like to think that Joyce's efforts are a useful supplement); my point is simply that we should be clear what Finlay does and does not accomplish. Even if his case were entirely successful, far from showing that the error theory is in error, all he would have shown is that one argument in favor of the error theory is in error.

Finlay partially acknowledges this when he confesses that “it is open to an error theorist to propose a different presupposition [than “absolute authority”] as the culprit for the systematic falsity of moral discourse, and ... such an error theory would not be vulnerable to the arguments of this paper” [351]. The acknowledgment is only partial because he doesn't see that the error theorist need not identify *any* presupposition—having instead been pushed into skepticism through becoming disenchanted by the centuries of failure exhibited by all alternative programs, for example. As for the possibility that the error theorist might support an argument of the (1)+(2) format based upon a different presupposition to that targeted by Mackie, Finlay is content to note that “absolute authority is the most familiar and promising candidate for the error theory,” and, as for possible alternatives, he writes: “I suspect that scepticism about absolute authority lies behind many even if not all of them” [ibid.].

But Finlay is mistaken about this. Not only is he mistaken in thinking that “absolute authority” lies behind most versions of the Argument from Queerness, he is mistaken in maintaining that it lies even behind Mackie's original version or my own variant. Finlay quite misidentifies the “poisoned presupposition” in question.

“The Mackian error theory,” writes Finlay, “claims (a) that moral discourse presupposes the absolute authority of moral value, while maintaining (b) that all genuine value has only contingent authority, which exists only for those with the right psychological orientation towards the relevant standard or end” [ibid.]. It is (a) that Finlay opposes.

Upon first reading this I was surprised to see the argument articulated in terms of “*absolute* authority.” Mackie talks a great deal about “*objective* values” and “*objective* prescriptions,” but hardly ever about “*absolute*” values or prescriptions. The term “absolute” can be used either in the vernacular or as a term of art. As a term of art it is associated with

the metaethical notion of *absolutism*, standing opposed to *relativism*. Relativism holds that moral claims contain an essential indexical element, such that the truth of any such claim requires relativization to some individual or group that can change from context to context. Absolutism denies this. The vernacular usage need have little or nothing to do with relativism. An ordinary speaker might claim that the tsunami was “absolutely enormous,” that Kate’s assertion was “an absolute lie,” that a monarch has “absolute authority,” and so on. These uses do not stand literally opposed to relativism. (The tsunami, no matter how large, is enormous only relative to a comparison class.) Thus, it is important that when we see the term “absolute”—even in the works of metaethicists—we do not leap to the assumption that a particular metaethical theory is under discussion.

The absolutism-versus-relativism dichotomy is orthogonal to the objectivity-versus-non-objectivity dichotomy. The enormity of a tsunami may be a relative matter (big relative to ordinary waves, small compared to a supernova), but there is nothing less-than-entirely-objective about the scale. Similarly, if the price of fish for everyone is fixed by a tyrant then there is something less-than-objective about that value (the truth of “The fish is worth \$5” will be determined solely by someone’s whim), yet this would not be an instance of *relativism*; there is no indexical that must be consulted—something that alters from context to context—in order to determine the truth value of “The fish is worth \$5.”

Thus, when Mackie claims that morality presupposes “objective values” and “objective prescriptions,” though precisely what he does mean is less than immediately transparent, he must not be interpreted as claiming that morality presupposes “absolute values” and “absolute prescriptions.” If that is what he meant to say, we can be sure that that is what he would have said. Let us attend to what Mackie did say.

My thesis that there are no objective values is specifically the denial that any such categorically imperative element is objectively valid. [1977: 29]

A hypothetical imperative is a piece of advice on how the subject may better satisfy his or her ends. “Don’t kill babies” might be uttered as a hypothetical imperative if it is offered as advice on how to avoid risking severe penalty (which the addressee greatly desires to avoid). Usually, however, “Don’t kill babies” will not be intended in that cold spirit; it will not be a piece of advice at all—rather, it will be uttered as a categorical imperative. Since the applicability of a categorical imperative does not depend on the addressee’s ends, it is tempting to refer to it as an “absolute” command. And indeed it does have a certain kind of absoluteness, inasmuch as there is a certain relation (i.e., between the action and the satisfaction of the subject’s ends) that is irrelevant to a categorical imperative. But this is not to say that categorical imperatives make sense only within the framework of metaethical absolutism, for they may be relativistic. Philippa Foot [1972] showed us how imperatives of etiquette are categorical. The command “Don’t speak with your mouth full” need not be a piece of advice (we are not forced to retract it when we discover that the subject doesn’t care a fig for politeness); and thus, since the relation of the advice to the addressee’s ends is irrelevant, this rule of etiquette has an absolute quality. But it is not an instance of *absolutism*. The speaker and audience may be well aware that “You ought not speak with your mouth full” is true relative to Western culture but false relative to some other culture (or, if you

prefer, that the imperative is “valid” relative to one culture but not valid relative to another⁴). The same options must be allowed for moral imperatives. When Mackie goes on to say “The objective values which I am denying would be action-directing *absolutely*” (and this is virtually the only instance of Mackie using the word “absolute” in this key chapter), he is best interpreted as saying that morality has absolute authority in much the same way that a monarch may have absolute authority: making demands that are inescapable, incontestable, non-negotiable. This is very different from declaring “The objective values which I am denying would be non-relativistic.”

Observe that Mackie nowhere denies the validity of categorical imperatives *per se*; he is denying the claim that categorical imperatives are *objectively* valid. What sense can we make of this distinction between “subjective categorical imperatives”—with which Mackie apparently has no beef—and “objective categorical imperatives”—the failure of which he supposes sinks morality? He immediately tries to explain: In any argument that issues a categorical imperative, he says, there must be some input (either a premise or the form of argument) “whose authority or cogency is not objective, but is constituted by our choosing or deciding to think in a certain way” [1977: 29]. Thus categorical imperatives are just fine, thinks Mackie, so long as they know their place: based on “our choosing or deciding to think in a certain way” rather than “part of the fabric of the world” [1977: 16].

The point is made somewhat clearer in Mackie’s later discussion of reasons. Imperatives connect to reasons claims: First, valid imperatives provide warrant for “ought”-claims; second, “X ought to ϕ ” is thought by Mackie to be equivalent to “There is a reason for X’s ϕ ing” [1977: 73-74]. A categorical imperative “Do ϕ ” addressed to X would thus imply a reason for X to ϕ that does not depend on ϕ ing promising to satisfy any of X’s ends. If Mackie were out to deny all categorical imperatives, he would presumably also deny the legitimacy of any such reasons claims. And yet he doesn’t do this at all. Suppose Amy comes upon a stranger writhing in agony and there is some action that she could perform to provide relief. “[I]t would be natural,” Mackie says, to claim that the stranger’s sufferings “constitute some reason” for Amy to act—a reason that does not depend in any way on Amy’s desires to help (or desires for the things that her helping might promote) [1977: 78-9]. Mackie explicitly admits these “desire-transcendent reasons,” just as he admits categorical imperatives. What he is adamant about, however, is that such reasons talk is made legitimate only by the presence of an *institution*: What allows the transition from “There is a stranger writhing in agony before Amy” to “Amy has a reason to help” is a cluster of institutional facts, not brute facts. Institutions have rules of conduct which guide the behavior and speech of adherents, and transgressions of which are condemned. Importantly, such requirements “are constituted by human thought, behaviour, feelings, and attitudes” [1977: 81], and thus any such requirements are, in a central sense, *non-objective*.

We are now in a position to restate Mackie’s argument more accurately: The Mackian error theoretic argument claims (a*) that moral discourse presupposes non-institutional desire-transcendent reasons and non-institutional categorical imperatives, while maintaining (b*) that all genuine desire-transcendent reasons are institutional and all genuine categorical imperatives are institutional.

⁴ Mackie uses the word “valid” in the quoted passage, I think, as a quality that stands to imperatives as truth stands to propositions. Had Mackie been discussing an “ought”-claim at this point (“You ought to ϕ ”) then he would have denied that such claims are ever objectively *true*. Since, however, he is discussing imperatives (“Do ϕ ”) the question of *truth* does not arise and he instead denies that such claims are ever “objectively valid.”

I find it very hard to assess the extent to which Finlay's several arguments against the aforementioned premise (a) "Moral discourse presupposes the absolute authority of moral value" also serve against the premise (a*). I am inclined to think that there must be some degree of misfit, if only because I cannot recognize myself in many of the views Finlay attributes to "Joyce (2001)." Rather than attempt to untangle and refute Finlay's arguments on a point by point basis, let me try to clarify certain key points of the Mackie/Joyce argument, with an eye to defending (a*).

It is common ground between Finlay, Mackie, and myself that "certainly moral judgments are characterized by categoricity in *some* sense; we do not explicitly condition our moral assertions on agents' desires or concerns, for example" [349]. Despite Finlay's grudging tone ("...in *some* sense") it turns out that he is with Mackie and me all the way here, granting that "it is essential to moral judgement as such" that it has categorical force [354]; moral commands are conditioned on the desires and concerns of those to whom they are addressed neither explicitly nor implicitly. Given this agreement, then, the dispute over (a*) turns entirely on whether moral discourse presupposes that the categoricity in question is *non-institutional* (which I am treating as a reasonable take on what Mackie means by "objective" in this context).

Why do I think that the categorical imperatives of morality are (conceptually) non-institutional? Let us reflect first on *institutional* categorical imperatives. Foot gave us the example of etiquette, while (as Finlay notes) Mackie discusses the rules of chess, and I appeal to the rules of gladiatorial combat. These are systems of rules that are created by individuals or groups and which can be "applied" to a someone regardless of whether that someone is interested or wants it. That the rules may be applied irrespective of the subject's interests may just be another rule of the institution.

In *The Myth of Morality* [2001], I tried hard just to motivate in the reader a sense of unease in thinking that morality might be like this. I discussed Plato's example of Gyges with his ring of invisibility, who sought to satisfy his basest passions by stealing and raping at will. Of course, there are all sorts of prudential reasons that one might give a person as to why he shouldn't act like this: He'll end up feeling disconnected from his community, people won't truly love him, he'll feel guilty, he'll get bored, etc., etc. But we are allowed to stipulate the example such that none of these speak to Gyges' true ends—he is simply depraved and doesn't care and enjoys his degenerate behavior. Because our moral framework is categorical we can carry on condemning Gyges—we can carry on legitimately saying "Gyges, you ought not do that!" till we're blue in the face. But if our utterances are merely a verbal output that has been validated by an institution of our own creation, then it all begins to sound rather shrill. We can picture Gyges pausing in some sadistic undertaking, acknowledging to us that, yes, we are speaking legitimately when we assert that he morally ought not act in this way, and then shrugging and carrying on. And if our categorical imperatives are nothing more than an institutionally-backed way of speaking, why should he do otherwise?

It should be stressed that I am not requiring that our moral condemnations should have the force to move Gyges to stop. Maybe his very depravity blocks his capacity to respond to reasonable argument. My point is rather that we in fact have no reasonable argument to offer him. If categorical moral considerations enjoy only *institutional* backing, then they do not represent reasonable arguments for Gyges (unless by "reasonable" we mean "moral"; but that gets us nowhere). Mackie allows that we can even truly say "Gyges, the suffering of your victims constitutes a reason for you to stop!"—but, again, once Gyges realizes that this reason

is the product of *institutional* facts—an institution, moreover, to which he does not personally subscribe—then not only will he remain unmoved but it is hard to logically fault him for this.

In Joyce [2001] I tried to pump the intuition that our unease with this scenario leaves us groping for *non-institutional* categorical imperatives and *non-institutional* reasons. But what on earth would *they* be like? And the answer is that I really don't know. I can honestly report that to the extent that I am familiar with moral judgment I have a very strong feeling that people want and require something more from morality, something objective and non-institutional. We want something that *really* binds Gyges, such that he is making a mistake *by his own standards*, even if he cannot be brought to recognize this. I have come to suspect that my inadequacy in articulating this idea is possibly not because of any failure of imagination or eloquence. If I am right, morality may be imbued with a deeply mysterious kind of force—a kind of primitive feeling of “being bound by rules and ends” that resists explication. It is as if we have a pre-wired superstition, which—our being the intelligent beings that we are—we cannot quite admit as such (and instead spend two thousand years trying desperately to defend). Perhaps Mackie and I fumble to dissect something that by its very nature cannot be brought into the light to be picked over by philosophical scrutiny.

Finlay claims that Mackie and I explicate this kind of inescapable authority in terms of persons being bound by morality “such that, regardless of their desires and concerns, they cannot disregard it without irrationality” [350]. This gives the wrong impression. In relation to the argument that Finlay is opposing, I have never asserted it to be a presupposition of ordinary moral thought that moral transgressors are irrational. I would rather attribute to ordinary thought an extremely inchoate presupposition, using a term like “practical oomph” [Joyce 2006] and possibly providing a bunch of examples like Gyges in order to try and capture the “feel” of moral authority. (And I will be careful to point out that *inchoate* does not imply *dispensable*.) With the appropriate accompanying explanations, I am willing to attribute to ordinary moral thought a presupposition about non-institutional desire-transcendent reasons and non-institutional categorical imperatives (though of course I don't claim that ordinary speakers think in these terms!). I do not maintain, however, that anything about *irrationality* necessarily follows from this; rather, the idea that the moral transgressor is irrational has had in my work the status of *a hypothesis*—a *prima facie* plausible but ultimately rejected hypothesis—about how an inchoate presupposition *might* be unpacked in precise and substantive terms. (And, as far as I can see, the claim that immorality presupposes irrationality plays no role in Mackie's work.)

Although I have no wish to rehearse my argument against moral rationalism as a conceptual claim, something needs to be said to explain some of Finlay's confusions about my intentions. As we have seen, he seems to interpret me (and Mackie) as arguing basically as follows (“F” for *Finlay*):

- F1) Morality is conceptually nonrelativistic.
- F2) In fact, a nonrelativistic morality is indefensible.
- ∴ The moral error theory is established.

Focusing on (F1), Finlay finds many places in *The Myth of Morality* where I appear to be arguing firmly (but, he thinks, unsuccessfully) in its favor. But I do not support the above argument; my arguments against moral relativism, upon which Finlay expends so much

critical attention, play a quite different role in the dialectic. Rather, as I indicated above, the argument starts out like this (“J” for *Joyce*):

J1) Morality conceptually involves non-institutional categorical imperatives.

I then ask what sense can be made of such imperatives, and whether they might be defensible. As a hypothesis, one might entertain:

H) Moral non-institutional categorical imperatives are rational requirements.

I then develop an account of practical rationality, following closely but critically in Michael Smith’s footsteps, coming to the conclusion:

i) Rational requirements are relativistic (in a certain way).

I then argue:

ii) But moral requirements are nonrelativistic (in that way), hence

iii) moral requirements cannot be rational requirements.

At this point I reject hypothesis (H), and, in the absence of any other plausible candidate for defending non-institutional categorical imperatives, feel justified in declaring:

J2) In fact, non-institutional categorical imperatives are indefensible.

∴ The moral error theory is established.

I won’t attempt to make that convincing; I’m just trying to clarify the structure of the argument in light of Finlay’s critique. One certainly will find me attacking moral relativism at some length in Joyce [2001], but this is in the service of supporting (ii), which is directed at sinking (H), the rebuttal of which contributes support to (J2). Should my arguments against moral relativism fail (and I dare say they might), there may well be other reasons to doubt (H) and other sources of support for (J2).⁵ (In fact, it would seem that Finlay will happily grant me (J2).)

Since Finlay misconstrues my arguments against moral relativism as aiming to provide support for (F1), he thinks that if a kind of moral relativism can be shown to be viable, then the entire master argument for the moral error theory collapses. In Section 2 of his paper, he locates multiple sources of evidence that I have supposedly put forward in support of the first premise of the error theoretic argument, and then proceeds to show how weak this evidence is. His contention is that the error theorist fails to prove his case; all this evidence can be explained just as well by the hypothesis that moral discourse presupposes a nonabsolute account of moral value as by the hypothesis that it presupposes the absolute authority of moral value. Given that Finlay declines to lay his cards on the table regarding just what kind of nonabsolutism he has in mind, it is mighty hard to compare hypotheses. In any case, I

⁵ In fact, in my [2001: 67-8, 120-1] I do deploy a different argument against (H) pertaining to the normative *content* of rationality versus our entrenched desiderata concerning the content of morality. Given the account of practical rationality I endorse, there appears to be a content mismatch sufficient to sink (H).

shan't go through these pieces of evidence in turn, since the points to rebuff and misunderstandings to clear up are many and complex. Suffice it to say that since the first premise of my and Mackie's argument is not the one that Finlay attributes to us, and since on those occasions that I do defend the premise that Finlay attributes to us it is in the service of a different (dispensable) argument, much of this debate misses its mark.

What does become clear, however, is that on nearly all occasions that Finlay attempts to show the nonabsolutist hypothesis accommodating some evidential datum, to make the explanation plausible he is forced to imagine the speaker unaware of the nonabsolutist quality of her conceptual scheme, or suppose this element tacit in her speech and thought. And while of course I agree that people are not always the best judges of what they are talking about, and that the presence of relativizing variables can be tacit, there is something fishy in the recurring need for Finlay to pull this move. The telling evidence is not whether people are thinking in absolute or nonabsolute terms when they are engaged in moral thought, but what they make of the nonabsolute variable when it is made explicit. Finlay correctly imagines how weird it would be if the All Black captain, in the midst of the rugby huddle, kept relativizing his advice with "...if you want to win the game." Yes, that would be weird and vaguely irritating. But, I submit, it would be a very different kind of weirdness if the judge at the Nuremberg trials kept relativizing his condemnation of the war criminals with the suffix "...by our moral standards." This would not just be weird and irritating; it would be *scandalous*; there would be protests. The All Black captain would be committing the infelicity of redundancy; he would be saying too much. The relativist Nuremberg judge, by contrast, will be interpreted not as adding something unnecessary, but as revealing himself, in adding the suffix, to be saying too little. The displeased audience will want the judge not merely to suppress that suffix in conversation; they will want him to *eliminate* it.

Part of what would be scandalous about such an explicitly relativistic judge is that he would be locating the locus of our moral condemnation in the wrong place. Were he to say to Hermann Göring "What you did was odious!" then he would clearly be talking about Göring, his actions, and his victims. But when he says to Göring "What you did was odious *from our point of view*," then we, the audience, have our attention turned also towards ourselves, in saying which I mean that both our emotive response and our sense of the justificatory source for our feelings shift to encompass ourselves. We would still hate and despise Göring and his actions, of course, but when it is made plain that this condemnation is (only) "by our moral standards" then our feelings may be affected by a discomfiting awareness of the *contingency* of our judgment.⁶

We can find more mundane examples of much the same phenomenon in etiquette. When one hears someone sniffing loudly and wetly in public, this can be rather disgusting. When one observes that the perpetrator is Japanese, and reminds oneself that such behavior is not considered such a breach of etiquette in that culture, then though it is not the case that one's disgust immediately evaporates (or even lessens), certainly *something* shifts in one's attitude. It is as if one no longer *endorses* the emotion in the same way. One's attention has moved

⁶ As far as I know, nobody called Göring "odious" at the Nuremberg trials. But the crimes against humanity of which he was convicted were subsequently formally defined by the Rome Statute of the International Criminal Court (1998) as (inter alia) "particularly odious offenses." (In the London Charter of the International Military Tribunal (1945), which created the legal basis of the Nuremberg trials, "inhuman" was the moral term of choice.)

from a simple “That’s disgusting!” to something broader and more complex—a questioning of one’s own perspective, an awareness of otherness, a consciousness of contingency.

In short, our emotional response towards someone who has transgressed against a norm which we judge he was objectively bound to follow is extremely different from our emotional response to someone who has transgressed against a norm that we endorse but recognize has only local applicability, and differs yet again from our emotional response to someone who has harmed himself (in frustrating the satisfaction of his ends). In all three cases, the surface form of our utterance may be “He ought not to have done that,” but this linguistic correspondence veils enormous differences. It is not that our negative feelings towards Göring would necessarily be weaker when they flow from a nonabsolute judgment, but Finlay seriously underestimates, I think, the extent to which they would be influenced. How much sounder we might sleep at night were we confident that we hanged the criminal because he did something *objectively* wrong, as opposed to acknowledging that we hanged him because he did something that *we found* wrong.

Finlay’s preferred view is that the judge’s relativizing suffix can be left tacit when the standards in question are shared by his audience. And when those standards are *not* shared, then by suppressing the suffix the judge can perform a rhetorical trick of expressing an expectation or demand that the audience *will* share those standards [357; see also Finlay 2004, 2005]. But this doesn’t explain the outrage that the audience will feel when the suffix is explicitly added. If Finlay’s analysis were correct, then by adding the suffix the judge causes his condemnation to lose its rhetorical flourish. In that case, the judge should be able to avoid scandalizing his audience by explicitly saying what the rhetoric aims to accomplish: “What you did was odious, from our point of view. And, moreover, I demand that people subscribe to our point of view.” This perhaps sounds *a little* better, but only, I think, by encouraging an illicit reading. What the judge *cannot* say is: “...I demand that people subscribe to our point of view *because it’s the correct point of view.*” To allow this would be to give up on the nonabsolutist analysis of the judge’s condemnation. Rather, the judge is simply demanding subscription. Now, the audience is likely to approve in general of the judge’s demanding subscription to moral standards. They are in the presence of someone (Göring) who, we are imagining, cares not a jot for morality, so saying to him “Subscribe to morality!” seems like a reasonable demand (though putting it in these terms manifests the usual philosopher’s clunkiness). But note that this is in fact *not* what the judge would be saying; he would be demanding “Subscribe to *our* morality!”—and now we are back where we started. The audience will balk: “No, we’re not just telling Göring to subscribe to *our* morality; we’re demanding that he subscribes to *the real* morality.” By comparison, the audience will *not* balk if instead of relativizing matters to a particular moral standard, the judge were explicitly to eschew all such relativization: “What you did was odious, *irrespective of one’s moral standards*” [see Olson 2010]. It would be astonishing if adding “...irrespective of standards” raises no eyebrows, but adding “...relative to our standards” prompts outrage, while in fact it is the latter kind of suffix that is *really* in play tacitly. And yet this is what Finlay would have us believe. It is, in short, implausible that when we morally condemn we leave off the relativizing suffix solely in order to achieve rhetorical effect; it is much more plausible that we add no relativizing suffix because there is none.

John Burgess points out that in Ambrose Bierce’s work of satirical humor, *The Devil’s Dictionary*, the following entry appears:

MORAL, *adj.*: Conforming to the local and mutable standard of right.⁷

“This can’t be an accurate transcription of the literal sense of the word,” says Burgess, “else how could it amuse? ... The wit of the ‘definition’ lies in its dramatic revelation of the gap between theory and the practice of Bierce’s Victorian contemporaries” [Burgess 2007: 429]. Later, Burgess stands by the principle, as do I, that “a definition that strikes people as a joke is unlikely to be a correct analysis” [439].

Many philosophers have given up on trying to make sense of moral objectivity because, like Burgess and myself, they have come to see that this “fugitive thought” (as Foot called it) is unlikely to be found. It is perfectly reasonable that forms of nonobjective revisionism should be explored, but the claim that moral speakers have never been much interested in moral objectivity in the first place is one that I find deeply implausible. Metaethicists who make this latter claim often betray their own declared position by undertaking every convoluted and distortion possible to squeeze every last drop of ersatz objectivity from their favored theory. What metaethicist has ever turned down a scrap of objective authority the accommodation of which could be claimed for his or her theory? If morality had never been concerned with objectivity in the first place, why this seemingly compulsory ritual that every metaethicist embarks upon of emphasizing all elements of his or her theory that lean in the direction of objective authority and downplaying all elements that lean away?

Suppose I am correct that speakers really have been “much interested in moral objectivity.” Finlay is willing to concede as much for the sake of argument (though he continues to refer instead to “absolutism”). In his final section he argues that “even if such an assumption is ubiquitous, we still have good reasons to conclude that it does not contaminate the meaning or truth-conditions of moral discourse, and hence that its falsity does not make moral judgement systematically false” [348]. So Finlay changes tack from trying to show that the error theorist has failed to make his case to trying to show “that the error theory is actually false” [360].

This touted killing-blow argument takes a while to arrive. After several diversions,⁸ the meat of Finlay’s case is finally laid out over three pages (i.e., pp. 365-7). By this stage he has usefully laid out a few contrast cases. On the one hand, we have examples where an error theory seems clearly correct: witches and phlogiston. On the other, we have examples where, although in the past we have held seriously false beliefs about the topic, an error theory seems wrong: motion and water. In the latter cases, our erroneous beliefs can be extirpated without bankrupting the subject matter; in the former cases, the erroneous beliefs “infect” the whole concept. Finlay is, at this point in the proceedings, conceding for the sake of argument that we have ubiquitously held erroneous beliefs about morality. So, as I put it in Joyce [2001: 96]: “The question we must ask is: ‘Is moral discourse more like talk of witches or more like talk of motion?’—and the answer is that it is more like talk of witches.” Finlay agrees upon the question but disagrees with my answer.

⁷ Other entries include: “LAWYER, *n.* One skilled in circumvention of the law” and “DUTY, *n.* That which sternly impels us in the direction of profit, along the line of desire.”

⁸ First, Finlay indicates that some form of “content-externalism” will sink the error theory, then, just as the case is developing, announces that he thinks that content-externalism is implausible for morality and so drops it. Second, he expounds arguments against moral fictionalism, which is, though a thesis that I have advocated alongside and in response to the error theory, entirely dispensable to the case for the latter. If one is going to criticize my case for the error theory, then let it be assumed, *arguendo*, that I am agnostic between fictionalism and eliminativism.

He alights upon my claim that the matter turns upon what “the point” of the discourse in question is. If moral concept M contains erroneous claim E, and transparently eliminating E would leave concept M*, then can we *use* M* in all or most of the ways that we heretofore used M? If not, then we have strong ground for thinking that M *essentially* presupposes E. My claim, then, is that a normative discourse for which categorical imperatives were transparently non-institutional (and desire-transcendent reasons transparently non-institutional) could not be used in many of the ways we use moral discourse, and therefore would not *be* a moral discourse. (It would at best be a schmoral discourse. [See Joyce 2007].) This is ultimately an empirical claim, and it has not been tested. But the kind of thoughts I was attempting to pump above, concerning how we feel about Gyges or about the relativist judge at Nuremberg, are designed to give us a premonitory hunch about how those facts might pan out.

We will not get far until we have a decent handle on to what uses we do put morality, and this too is an empirical question. (I have offered speculations in the past, but I bow to the tribunal of our best evidence.) Finlay presumes that I intend to invoke only the “*referential* point” of morality [364], but I must confess that I don’t quite know what this means. I agree that there may be *many* “points” to moral discourse; my position is that many of these make sense only if we take ourselves to be referring to non-institutional values and prescriptions. The point—or better, *a* point—of punishing Göring is to gain satisfaction in seeing him get what we think he deserves—where the kind of desert in question is embedded in a meaning-granting conceptual scheme according to which Göring’s actions were objectively wrong, not merely wrong from our perspective.

Finlay claims that just as “the absolutist about motion makes (what look like) substantially the same first-order motion judgements as the rest of us [who are sensible motion relativists], ... the absolutist about morality makes (what look like) substantially the same first-order moral judgements as the rest of us” [365]. But this is precisely the claim about which I am doubtful. And while it may be true that I haven’t said enough to ground this doubt persuasively, by the same token Finlay hasn’t said enough simply to help himself to the claim. (And relative to the ambition of showing that the error theory “is actually false,” his argument requires that he dispels this reasonable doubt.) He goes on:

Assuming there is no genuine absolute motion, or genuine absolute moral properties, the absolutist’s judgements could not be responsive to these fictional properties. Rather, his judgements about motion are responsive to his sensitivity to motion relative to particular frameworks, and his judgements about moral wrongness are responsive to his sensitivity to the relation of actions to certain moral standards or ends.
[365]

On one reading, this is probably platitudinously true. No matter what moral judgment Abe makes, we can probably locate a framework of “standards and ends” relative to which his judgment is true. But this isn’t really what Finlay wants, for the aim is not to ensure that Abe’s moral judgments must come out as *true*. Even if Abe is a decent person, Finlay will want to allow that his moral judgments (just like his judgments of motion) are on occasion false. I suppose this is why Finlay doesn’t claim that the absolutist’s moral judgments are always “responsive to the relation of actions to certain moral standards or ends,” but rather adds an extra step: they are “responsive to *his sensitivity to*” this relation. Even when the absolutist makes a mistake—and thus fails to be responsive to the relevant relation of actions to a certain framework of ends—he is still responsive *to his sensitivity for* tracking a real,

albeit nonabsolute, realm of fact. Why, then, should we not just see the absolutist as employing the nonabsolutist moral concepts all along?

I suspect that lurking in the background of Finlay's thinking is a quantifier-shift fallacy. It may be true that for every absolutist judgment Abe makes, there exists some nonabsolute property (pertaining to a framework of ends relative to which his judgments may be assessed) to which he is manifesting a sensitivity, but it does not follow that there exists some nonabsolute property (etc.) to which Abe is manifesting sensitivity every time he makes an absolutist moral judgment. But it is the latter claim, the $\exists x \forall y$ proposition, of which Finlay needs to convince us. He wants to maintain that there is some nonabsolute standard relative to which a person's absolutist moral judgments can be uniformly assessed. Now, it is possible that this claim is true, but Finlay cannot get it for free simply on the basis of the platitude that people make moral judgments and there will nearly always be *some* instantiated relational property to which each judgment can be interpreted as a response.

Finlay sees the obvious flaw in this whole line of thinking, and tries valiantly to fix it. The flaw is that even for those topics for which an error theory seems obviously correct (witches and phlogiston), when people in the past made these judgments there was probably some kind of instantiated property to which their judgments could be interpreted as a response. In the case of witches, I once suggested that the instantiated property (if there is one) might be that of "playing a certain disruptive role in the patriarchal society" [Joyce 2001: 96]; Finlay prefers: "a complicated disjunctive property including, for example, the property of being a woman whose enemies have suffered illness and misfortune" [366]. But locating such a property wouldn't lead us to overrule our witch skepticism. So Finlay's attempt to drive a wedge between witches/phlogiston and motion/water, such that morality clearly falls on the latter side, by his own admission initially fails.

His "fix" is to distinguish between judgments that are responsive to (what we take to be) *evidence* of their instantiation, and judgments that are responsive to that which is (we take it) *constitutive* of their instantiation. Witches and phlogiston are supposed to be of the former type, and motion (and presumably water) of the second type. With his new wedge driven in, Finlay again expresses confidence that morality falls on the latter side.

Serious problems emerge, however. In the witch case, we are invited to observe that

competent users of the concept *witch* will have no trouble conceiving of a situation where the standard evidence of witchcraft obtains, but the accused is not a witch, or of a situation where none of the standard evidence obtains (no wart, no black hat, no sick enemies) but in which a woman is a witch nonetheless. [366]

Fair enough. But I fail to see a distinction with morality in this respect. Let's just re-run the quote again, but this time for morality:

Competent users of the concept *moral wrongness* will have no trouble conceiving of a situation where the standard evidence of moral wrongness obtains, but the action (say) is not morally wrong, or of a situation where none of the standard evidence obtains (no apparent suffering, no hurtful intentions) but in which the action is wrong nonetheless.

There seems nothing dodgy about that to me. But Finlay thinks otherwise, arguing that it is "not nearly as plausible" that moral judgments "could be sensitive to what a person takes to

be merely evidence for the realization of” absolute properties (which are in fact nonexistent) [366]. In explaining this, he makes the curious claim:

Consider that the “evidence” in each case [both motion and morality] is the relational property that the relational theory identifies with morality or motion. [366].

But why is *that* the evidence? In the above paraphrase for the moral case, I listed as “standard evidence” such things as apparent suffering and hurtful intentions. (That was just a flavor of a real answer—but it will do for present purposes.) It seems bizarre to me that instead it is *the relational property* in question that Finlay counts as evidence. This table may help clarify matters.

	The evidence	The (possibly) instantiated property	The uninstantiated property
WITCHES	e.g., testimony about a woman’s suspicious behavior, accusations from upstanding citizens, coerced confessions, etc.	“a complicated disjunctive property including, e.g., the property of being a woman whose enemies have suffered illness and misfortune” [366]	being a woman who can cast spells and has entered into a pact with the devil
MORALITY	e.g., the presence of suffering and selfish intentions, etc.	nonabsolute relational property pertaining to the action’s bearing on the speaker’s ends	absolute moral property

In the witch case, Finlay clearly distinguishes these three categories (though for the first box he instead mentions pointy hats and warts⁹), but in the morality case, the first column seems to pass unnoticed and instead the nonabsolute relational property (2nd column) is treated as the “evidence.” This is perplexing. I cannot see why Finlay supposes that the error theorist might be the least tempted to impute to ordinary moral speakers the tendency to treat the instantiated relational property *as evidence for* an absolutist property (which is in fact uninstantiated). Rather, the error theorist thinks that ordinary speakers treat things like suffering and selfish intentions as evidence for the presence of an absolutist property (which is in fact uninstantiated). Perhaps these things like suffering and selfish intentions are in fact reliable indicators of the presence of some instantiated relational property; but the error theorist has reasons for thinking that this relational property will not transparently serve in place of the uninstantiated absolutist property.

I do not see that things stand any differently in the witch case. The error theorist about witches (i.e., us) thinks that superstitious speakers in the past treated things like testimony about a woman’s suspicious behavior, accusations from upstanding citizens, coerced confessions, and so forth, as evidence of the property of witch-hood (which is in fact uninstantiated). Perhaps these kinds of things were in fact reliable indicators of the presence of some “complicated disjunctive property” that the accused women really did instantiate; but the error theorist (i.e., us) has reasons for thinking that this complicated disjunctive property will not transparently serve in place of the uninstantiated property of being a witch.

⁹ I take it that he wasn’t being entirely serious at this point.

If Finlay insists on making the odd claim that error theorists think that ordinary people treat the nonabsolute relational property as evidence of the presence of an absolute moral property, one can just counter with the no less odd claim that ordinary people treat the complicated disjunctive property (which all women accused of being witches in fact instantiate) as evidence of being a witch. In either case, no obvious asymmetry has been exposed.

It should be stressed that this is really the pivotal moment in Finlay's argument. The lengthy preliminaries, the careful distinctions, the twists and turns—all don't amount to too much, since at the point immediately prior to this he has admitted that "something has gone wrong here" with his argument; he has seen that the error theorist still has a wide open escape route. (In fact, it's not as if the error theorist has had to dodge and weave and then finally glimpses a way out; the "escape route" could have been described by the error theorist at the outset.) And yet at the crucial moment where Finlay is supposed to slam the door on that escape route, as far as I can see he has no well-structured or convincing argument to back up the rhetoric. I should emphasize that I do not take myself on this occasion to have said anything convincing about why the moral case must go with witches and phlogiston rather than motion and water; my intention is simply to refute an argument that purports to provide decisive support for morality going with motion and water.

However, it may be that in advocating these claims I am merely manifesting the blinkering influence of an upbringing in a "melting pot made up of immigrants from a wide variety of backgrounds" [368]—a peculiarly Australasian misfortune that has clouded my capacity to think clearly on these matters. I say this in response to the curious closing portion of Finlay's paper where he aims to analyze why the moral error theoretic stance is a "characteristically antipodean view" [347]. At the risk of being impolite, I think his comments are most charitably interpreted as a joke. If one were to entertain such wild ad hominem speculations, one could just as easily attribute to the antipodean error theorist a kind of *courage* to sincerely question deeply entrenched human beliefs in the face of overwhelming opposition: a courage born of the spirit of those plucky Anzacs confronting a deeply entrenched enemy in the hills of Gallipoli. One might with as much warrant claim that the antipodean error theorist's refusal to kowtow to conservative forces finds its origins in the tenacious self-sufficiency of the early pioneering families, in the fierce autonomy of Ned Kelly and Hone Heke. But oh what silly arguments those would be.

University of Sydney

References

- Bennett, Jonathan, 1995. *The Act Itself*, Oxford: Oxford University Press.
- Burgess, John P. 2007. Against Ethics, *Ethical Theory and Moral Practice* 10: 427-439.
- Doris, John 2002. *Lack of Character: Personality and Moral Behavior*, Cambridge: Cambridge University Press.
- Finlay, Stephen 2004. The Conversational Practicality of Value Judgement, *Journal of Ethics* 8: 205-23.
- Finlay, Stephen 2005. Value and Implicature, *Philosophers' Imprint* 5: 1-20.
- Finlay, Stephen 2008. The Error in the Error Theory, *Australasian Journal of Philosophy* 86: 347-369.

- Foot, Philippa 1972. Morality as a System of Hypothetical Imperatives, *Philosophical Review* 81: 305-16.
- Joyce, Richard 2001. *The Myth of Morality*, Cambridge: Cambridge University Press.
- Joyce, Richard 2006. *The Evolution of Morality*, MIT Press.
- Joyce, Richard 2007. Morality, Schmorality, in P. Bloomfield (ed.), *Morality and Self-Interest*, Oxford University Press: 51-75.
- Joyce, Richard 2009. Patterns of Objectification, in Richard Joyce and Simon Kirchin (eds.), *A World Without Values: Essays on John Mackie's Moral Error Theory*, Springer Press.
- Joyce, Richard and Simon Kirchin (eds.) 2009. *A World Without Values: Essays on John Mackie's Moral Error Theory*, Springer Press.
- Lewis, David [1989] 2000. Dispositional Theories of Value, in his *Papers in Ethics and Social Philosophy*, Cambridge: Cambridge University Press: 68-94.
- Libet, Benjamin 2004. *Mind Time: The Temporal Factor in Consciousness*, Cambridge MA: Harvard University Press.
- Olson, Jonas 2010. In Defence of Moral Error Theory, in M. Brady (ed.), *New Waves in Metaethics*, Palgrave Macmillan.
- Prinz, Jesse J. 2007. *The Emotional Construction of Morals*, Oxford: Oxford University Press.
- Roskies, Adina 2003. Are Ethical Judgments Intrinsically Motivational? Lessons from Acquired Sociopathy, *Philosophical Psychology* 16: 51-66.
- Shafer-Landau, Russ 2005. Error Theory and the Possibility of Normative Ethics, *Philosophical Issues* "Normativity" 15: 105-120.