A Semantic Paradox concerning Error Theory

STEPHEN HARROP

J. L. Mackie famously argued for a moral error theory—that is, the thesis that our statements concerning objective moral value are systematically false.

§1 offers a rephrasing of Mackie's thesis in more specifically anti-objectivist terms and explicates his motivations for that theory. Further, it offers a distinction between different types of error theory.

§2 advances the position that Mackie's view—and a type-1 error theory in general—implies a contradiction and is thus false. It also offers 4 possible responses the error theorist might take, and consider options 1–3 and their consequences.

§3 examines variations of option 4, mainly having to do with possible translations of the negation of a particular moral predication P.

§4 considers answers to possible responses to 4 that the error theorist might offer and gives reasons to think them insufficient.

§5 considers a possible reformulation of a 1st-order error theory in response to the points raised in §4, and provide arguments as to why it is untenable.

Stephen Harrop received a BA from Johns Hopkins University in 2016 with concentrations in mathematics and philosophy. He'll be returning to Hopkins in the fall to pursue an MA in philosophy. His philosophical interests are in early modern philosophy (especially Leibniz and Reid), medieval philosophy (especially Scotus, Aquinas, medieval logic) and metaphysics.

I. Exegesis and Definitions

J. L. Mackie denied that there is any such thing as an objective value:

> The denial of objective values [is] put forward not as the result of an analytic approach, but as an 'error theory', a theory that although most people in making moral judgements implicitly claim, among other things, to be pointing to something objectively prescriptive, these claims are all false. (35)

We can phrase the general error-theoretic thesis as:

> (E1) Error theory is the thesis that all predications of moral value are false.

The notion of a moral predication operative here is just the intuitive one—a moral predication is a statement of the form "X is A", where X is some entity and A is a predicate containing one or more terms with moral or morally evaluative significance. This is a somewhat over-broad definition, since there might be different types of moral value; dismissing them all categorically misses distinctions which might inure certain types of value to such dismissals. Consequently, we'll parse the definition given in E1 as the conjunction of two separate theses:

> (E1.1) All predications of objective moral value are false (type 1)

> (E1.2) All statements of predication of subjective moral value are false (type 2)

Mackie only seems to assent to the former; he does after all refer to his view as a species of moral subjectivism, so presumably statements where the moral value predication is merely one of a person-relative or subjective value will not always be false. Mackie doesn't offer a very concrete definition of what "objective" means in this context, so we'll understand "subjective values" to be "values which are agent- or mind-dependent," and "objective values" to be "values which are agent- or mind-independent."

Following Mackie's piece, we'll take error theory to be defined as given in E1.1. We should note that, while indeed E1 entails both

E1.1 and E1.2, it is not the case that either of these entails E1; nor does E1.1 entail E1.2 or vice versa.

In light of these distinctions, we'll take Mackie to be arguing against the view that there are such things as true predications of objective moral value to some action, conjunction of actions, state of affairs, and so on. His thesis is just that, if P is a moral predication, P is always false.

He justifies this thesis by means of two arguments: the argument from relativity and the argument from queerness. These arguments have no direct bearing upon my point, so I shall not examine them in detail.

The argument from relativity runs thus: "radical differences between first order moral judgements make it difficult to treat those judgements as apprehensions of objective truths" (Mackie 36).

The argument from queerness runs thus:

> If there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe. Correspondingly, if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else. (Ibid. 38)

I won't comment on the validity of these arguments; I put them forth here to show that Mackie's thesis isn't unmotivated. If we can defeat the thesis he claims to justify, though, what are we to make of these arguments? We'll briefly return to that question at the end of §5.

II. The Principle Argument

My thesis is that Mackie's view, given certain other relatively uncontroversial positions, is a position whose affirmation entails a contradiction. This will be true of any other type-1 error theory as well. To be clear, it is not a direct claim of self-referential incoherence, as is often made of the principle of verification—e.g., "if only that which is verifiable is meaningful, how is the principle of verification meaningful by its own standards?" Rather, I mean the following: if a

type 1 error theory is true, then a contradiction is true. These are, I recognize, rather strong claims, and I'll do the best I can to motivate them in the following pages.

The central argument is pretty easy to sum up: assume error theory is true. If all moral predications are false, and P is some moral predication, then P is false. If we think that at least some negations of moral predications are themselves moral predications, then we also have that ¬P is false, and thus that ¬¬P is true. But ¬¬P=P, so P is true. So P and ¬P. Therefore, error theory is false.

But we need to spell this out in a little more detail. Let's represent the moral claim "murder is objectively wrong" as P, and the moral claim "murder is not objectively morally wrong" as ¬P. More formally, let *m* stand for "murder", and OW(m) stand for "action *m* is objectively morally wrong".[1] Then P=OW(m) and ¬P=¬OW(m).

There are two possible ways of translating ¬P: either as "murder is morally neutral" or "murder is morally good." For Mackie's purposes, though, either of these translations will do, as both claims are by hypothesis false. Let us take v(P) to be the truth-valuation of the sentence P, with v(P)=T if P is true and v(P)=F if P is false. So, by Mackie's hypothesis, we have both v(P)=F and v(¬P)=F. But if we accept the law of the excluded middle, either P or ¬P is false (i.e., either v(P)=T or v(¬P)=T). So assuming both Mackie's hypothesis and the law of the excluded middle, we have both v(¬P)=T and v(¬P)=F, which is a contradiction.

If the above argument is valid, the error theorist has, so far as I can see, four options:

> (1) He can reject the law of the excluded middle.

> (2) He can say that there are some true moral claims.

> (3) He can reject error theory and say that moral claims are not the sort of thing to which you can assign truth-values.

---

[1]Here, one can take one of two routes. Either one can take OW to be a compound predicate, or else one can take W to be a predicate and O a sentential operator. The particular interpretation doesn't do much to change the argument.

(4) He can reject my translation of ¬P.

It's doubtful that the error theorist will want to take option 1. Nor can he take option 2 and remain an error theorist. That leaves 3, that moral claims are not the sort of thing that can have truth value (much like semantic paradoxes), and 4, that my rendering of ¬P is wrong. Since 3 will also entail giving up error theory, this would not be an acceptable option. Therefore, 4 is the only real option available to the error theorist. 4 is the topic of the following section.

III. On Interpretations of ¬P

We'll now turn to the question of whether the offered interpretation of ¬P is either plausible or correct. Pretty immediately, the error theorist might justifiably ask: "wait, how does ¬P distribute into P? Doesn't that depend on the particular syntax of any given P-like predication?" This is a genuinely interesting, and not unmotivated, question, and so we'll try and answer it. In what follows, we'll take words like "wrong", "right", "bad", "good", and so forth to have their specifically moral meaning: morally wrong, morally right, and so forth. We begin by examining the following:

(P1) It is wrong to murder.

This statement is obviously a moral statement. Furthermore, the claim is that this is a rephrasing of

(P0) Murder is objectively wrong

This preserves all the relevant semantic structure of the original claim. I won't offer a full-blooded defense of this claim, but it has roughly the following motivation: suppose we use •A• to denote the propositional content of any sentence A. Then the claim becomes •P0• = •P1•. Consequently, any general claims that I'll make about P1 will carry over to P0. So P1 is also a moral predication in the relevant sense

The characterization "is a moral predication" is not quite as obvious in the case of its negation, in part because it is not clear just what part of the sentence is negated. For instance, we might say that

(¬P1.1) It is objectively wrong to not murder

is the correct translation of this operation. However, it seems implausible that this is the relevantly correct version of ¬P, as this version says that murder is morally required, not that it is not objectively wrong to murder. So to assume that this is the correct translation may beg the question against the error theorist. So we might think the following is a plausible interpretation of ¬P:

> (¬P1.2) It is not objectively wrong to murder.

This gives an opportunity to put into more explicit language the misgivings we have with these sorts of statements: given a moral predicate OW(x) or OR(x),[2] and an assertion connective ! (which just appends to the statement the phrase "it is the case that"), we can translate the propositions we've given so far as follows:

> (¬P1.3) "It is the case that it is objectively not wrong to murder" = !O¬W(m)

> (¬P1.4) "It is not the case that it is objectively wrong to murder" = ¬!OW(m)

> (¬P1.5) "It is the case that it is not objectively wrong to murder" = !(¬OW(m))

> (¬P1.6) "It is the case that it is objectively wrong to not murder" = !(OW(¬m))

We can now try to examine with a little more clarity which of these is making a moral statement and which is not. ¬P1.3 and ¬P1.6 clearly are making moral statements. Whether ¬P1.5 is a moral statement in all cases turns on whether we think that the negation of one of a set of bivalent moral predicates (wrong/right, good/bad, etc.) entails the other value. It seems to me that this is the case. If we think that moral predications can have truth value, then we can have a truth-valuation of a moral predication. For instance, if we have the moral predication OW(m) ("it is objectively wrong to murder") and we take a truth valuation of it, v(OW(m)), then pretty obviously either v(OW(m))=T or v(OW(m))=F . If v(OW(m))=T, then it follows

---

[2](OW() = "it is objectively wrong to," OR( ) = "it is objectively right to", x = some action)

immediately that $v(\neg OW(m))=F$. The converse is true as well; additionally, $v(\neg OW(m))=T$ entails that $v(OW(m))=F$. .So I hold we can conclude reasonably that ¬P1.5 is also making a moral predication.

We now turn to the more interesting case of ¬P1.4. In this case, we have the negation being applied onto the assertion operation, rather than on the moral predicate OW(m). This might initially seem consequential, since one often can derive a fallacy from an operation or quantifier. I don't think this works here, because $v(!OW(m))=T$ iff $v(OW(m))=T$. If we put this into natural English, this becomes more apparent: the proposition "it is the case that it is objectively wrong to murder" is correct only if the proposition "it is objectively wrong to murder" is true. Likewise, the proposition "it is objectively wrong to murder" is true only if the proposition "it is the case that it is objectively wrong to murder" is true. So the one can't be true without the other, which implies they're equivalent. So we get that ¬P1.4 is equivalent to ¬P1.5—both of which, incidentally, are for our purposes the same as ¬P1.2.

How can the error theorist—who presumably wishes to maintain that no negation of a moral statement entails the truth of a corresponding moral statement—respond? He might deny that ¬P1.4 is a moral predication. In other words, he might hold that, while ¬P1.3 and ¬P1.6 do express genuine moral predications, ¬P1.4 is completely compatible with a state of affairs in which there are no moral qualities, relations, and so forth. In that case, ¬P1.4 would simply express the truth that there is no fact about the world which makes it the case that it is objectively wrong to murder. In that case, ¬P1.5 expresses the same thing. They are ture only vacuously, and thus he can escape the charge of contradiction without any trouble.

IV. A New Hope for 4

Is there a response to this objection? I think there is. In brief, it is the following: if the error theorist wants to maintain that there's a translation of ¬P that isn't a moral predication, he's committed to the position that no proposition of the form ¬P1.4 is a moral predication. Otherwise, his theory still implies a contradiction. He must say, then, that propositions of the form "it is not the case that X is morally wrong" are necessarily not moral predications. At first

glance alone, this is an enormous burden of proof. Still worse, I will provide what I think are counterexamples to this proposition. First, however, we need some terminology.

Let's consider A, the set of all atomic moral predications—that is, predications like "X is wrong" or "X is good" for some action X, in the simplest possible description of moral rightness or wrongness.[3] If we think that A is closed under negation—in other words, that the negation of some atomic moral predication a is itself a member of the set of atomic moral predication—then it's immediately clear that $\neg a$ is itself a moral predication, subject to the error theorists contention of negative truth-value.

But suppose the error theorist denies that A is closed under negation as indeed he must. Then one of several things is the case.

First, it might be the case that the negation of no atomic moral predication is either itself an atomic moral predication or that it entails an atomic moral predication (or indeed any moral predication). This will henceforth be referred to as the disjoint hypothesis, or DH.

Second, it might be the case that the negation of only some atomic moral predications are themselves atomic moral predications. This will henceforth be referred to as the overlapping hypothesis, or OH.

Finally, it might be the case that either the overlapping or the disjoint hypothesis is true, but that A is closed under some joint negation/entailment operation: the negation $\neg a$ of some atomic moral predication a itself is not necessarily an atomic moral predication, but instead entails another atomic moral predication b. This will henceforth be referred to as the entailment hypothesis, or EH.

It should be pretty evident that neither OH nor EH will suffice for the error theorist's purposes. If OH is true, then some negations of moral statements are themselves moral statements, and we have the contradiction yet again. If EH is true, then the negation of a moral statement is not itself a moral statement, but

---

[3]Or, in the spirit of the greatest possible ecumenism, goodness or badness, or however one wishes to parse out the precise terminology of morals.

still entails the truth of some other moral statement. Since the error theorist is committed to the falsity of all moral statements, he then maintains a view that holds both that the entailed moral statement is true and that it is false, which is a contradiction. So the entailment hypothesis doesn't work either.

So only DH will serve to establish the error theorist's hypothesis. This is unfortunate for him, since it commits him to a view of the negation of moral propositions that is either necessarily true or necessarily false, which is a truly spectacular burden. If even one counterexample arises, his theory falls.

But need he worry? Are there such counterexamples? I think so. Consider the following:

> (P2) It is not the case that it is not permissible to murder.

This seems, on the face of it, to be a moral predication. It asserts that it's not true that it is impermissible to murder. Or, rather, there's a translation of this statement under which it is a moral predication. If the error theorist denies that this is a moral predication, he will have to motivate that denial ina fashon that is not ad-hoc. To that end, he might say that, since the above is the negation of the statement "it is the case that it is not permissible to murder", then to assert that this is indeed a moral predication is just to beg the question against him.

Does this response work? I'm inclined to think not. Let's put P2 into a more formal mode:

> (P2.f) $\neg(!(\neg P(m)))$

As we saw before, statements of the form $\neg(!OW(m))$ and $!(\neg OW(m))$ are equivalent. Thus, substituting in the moral predicate P, we have:

> $\neg(!(\neg P(m)))$ iff $(\neg(\neg(!P(m)))$ iff $(!P(m)))$

since double negation cancels. So it is in fact a moral statement, even if we accept that statements of the form given by $\neg P1.4$ and $\neg P1.5$ are not moral statements, and the charge of mere proof-by-assertion fails. And thus we get the counter example. The negation of P2 is:

(¬P2.1) It is not the case that it is not the case that
it is not permissible to murder.

which is equivalent to

(¬P2.2) It is the case that it is not permissible to
murder.

This certainly is a moral predication, so the counterexample holds.
So the disjoint hypothesis also fails to give the error theorist his
desired result; he is, colloquially, up the creek without a paddle yet
again. Even if he maintains 4—our translation of the negation of the
given moral predication P is faulty—he appears to have to embrace a
reading of ¬P that still entails the contradiction he sought to avoid.
And so the only route left is brute denial that the entailment ever
follows.

V. One Response from the Error Theorist Considered

But is there a way to avoid this entailment—and necessarily
avoid it—beyond the response we considered in §4, and without
falling prey to sheer despair and contrariness? Maybe so. One pos-
sibility is offered by Charles Pigden:

There is only one way out for the nihilist. He has
to deny the RD principles. 'Action X is not wrong',
does not entail that action X is right (in the sense
of morally permissible) nor does 'action X is not
right', entail that action X is wrong. But is this
bold and blunt assertion anything more than the
desperate response of the cornered nihilistic rat?
No, because (I think) it can be motivated. A entails
B if it cannot be the case that A is true and B false.
Or A entails B if there is no conceivable situation
(possible world) in which A is true and B false. Is
there a conceivable situation in which (¬B) is true
and (Br) false? Yes. The situation in which there are
no moral properties or relations, and specifically no

>properties of rightness, wrongness or obligatoriness
>which attach themselves to acts. (453)[4],[5]

If I were to offer a rephrasing of his argument, it would go roughly as follows: if statements of the form "Action X is not wrong" entail "Action X is right", then they must never fail to entail the consequent. But if the error theorist is correct, then it is in fact the case that there is some situation in which the antecedent is true and the consequent not: at that possible world in which there are no objective moral properties, duties, etc, which—or so will say the error theorist—is in fact the world we inhabit. So the denial of that entailment is not unmotivated, and the error theorist is out of the woods.

But this won't do. What this response gets us is that entailments of the form "'Action X is not wrong' entails 'Action X is right'" are not truth-preserving. That is, they're not true for every value of X, or every value of "Action X is/is not wrong". What it does not get us is that they are necessarily false, which is precisely what the error theorist would need in order to defend his position. Recalling the terminology of §4, what this at most gets us is the overlapping or entailment hypotheses. But these are far too weak for the error theorist's purposes. So this response is likewise insufficient.

The error theorist might try and defend a type 1 error theory which entails that there are no objective moral properties. Suppose I were to defend the following type 1 thesis:

>(E2.1) All statements of predication of objective
>moral value are necessarily false.

---

[4] It should be noted that Pigden is here responding against a position similar but not quite identical to my own. This is, however, not consequential at the moment, since I think his argument can be refashioned as an attempted response to mine. Additionally, I have good reason to be grateful to Pigden; much of the phrasing of §4 is straight out of his paper, and shone a light of concision and noetic clarity into what was otherwise a tricky and muddled subject and argument.

[5] "RD" here refers to the "reinforced doppelganger" principles mentioned earlier in Pigden's paper: They are as follows:
   (RD1) 'It is not the case that action X is wrong', entails 'action X is right'.
   (RD2) 'It is not the case that action X is right', entails 'action X is wrong'." (452)

If this thesis is true, it turns out that all entailments of the form "'Action X is not wrong' entails 'Action X is right'" are false. And so the type 1 error theory would be saved.

This, it seems to me, is a wildly over-committed theory. For one thing, it commits the error theorist to the position that, for instance, Plato's Form of The Good is necessarily non-instantiated at any world. More generally, it commits the error theorist to the view that objective moral properties are necessarily non-instantiated. This runs very deeply against our intuitions regarding modality, for surely it is prima facie possible that there is some world where an abstract object sufficiently like the Form of The Good is instantiated.

The error theorist may respond by saying that he does not trust our intuitions about objective moral properties anyway, since he thinks that all common-sense objective moral predications are simply flat-out wrong, and so to say "it's unintuitive" is something of a question-begging move. But this is a category mistake. Certainly it may the case that our intuitions regarding moral properties in this world are false; but that is not the intuition we're talking about. Rather, these are intuitions regarding possibility and necessity, about necessary instantiation and necessary non-instantiation. The intuitions that the error theorist denies are intuitions regarding the actual world; intuitions about possibility are entirely different and are not subject to the same range of limitations that our intuitions about what actually obtains carry.

And in any event, I think the counterexample I offered still holds. E2.1 seems to logically entail that all entailments of the form "'Action X is not wrong' entails 'Action X is right'" are necessarily false—that is, that there is not a single counterexample to it. But there clearly was one provided in §4. If that counterexample is valid, then we can conclude by simple *modus tollens* that E2.1 is false.

This leaves the error theorist back in the dilemma of having to choose between 1 and 2. Either he bites his lip and admits that the moral realist has something of a point, or else he throws up his hands and joins the non-cognitivists in saying that moral predications have no truth-value. But in neither instance can he retain his error theory. Or, I suppose, he could deny the law of the excluded middle is necessarily true and take this as one of the instantiations of its falsehood.

VI. Conclusion

In §2 I presented my central argument, that Mackie's claim is false—indeed, necessarily false—and I spent basically the entirety of §3–§5 defending this contention against various objections. But does this suffice to show that there are in fact objective values? Is a refutation of type 1 error theories sufficient to show that our claims about objective moral considerations are globally—or even locally—correct?

No. It may still be the case that the non-cognitivist is right. Indeed, as option 2 indicates, his position is entirely compatible with the arguments offered above. Rather, it means that we cannot systematically apply a single truth-value to moral claims—be it "false" or, indeed, "true"—without falling into a paradox not easily escaped. Whatever the correct theory of moral predications happens to be, it cannot not be given by cutting the Gordian knot of truth-valuations with the Myrtale of categorical falsity or truth.

Works Cited

Mackie, J. L. *Ethics: Inventing Right and Wrong.* Harmondsworth: Penguin, 1977.

Pigden, Charles R. "Nihilism, Nietzsche and the Doppelganger Problem." *Ethical Theory and Moral Practice* 10.5 (2007): 441–56.