

the moral truth

We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights...

Common-sense allows that talk about moral truths makes perfect sense. If you object to the United States' Declaration of Independence's assertion that it is a truth that 'all men' are 'endowed by their Creator with certain unalienable Rights', you are more likely to object that these rights are not unalienable or that they are not endowed by the Creator, or even that its wording ignores the fact that women have rights too, than that this is not the sort of thing which *could* be a truth. Whether it is a truth or not seems beside the point, anyway; the writers of the Declaration could just have well written, 'We hold it to be self-evident that all men are created equal, and also that it is self-evident that all men are endowed by their Creator with certain unalienable Rights,' save only that its cadence would lack some of the poetic resonance of the version which garnered Hancock's signature.

Yet famously, ethical noncognitivists have proclaimed that moral sentences *can't* be true or false – that, like 'Hooray!' or 'dammit!', they are not even the *kinds* of things to be true or false. Noncognitivism is sometimes even *defined* as the view that this is so, but even philosophers who define 'nonscognitivism' more broadly, as consistent with the idea that moral sentences may be true or false, have believed that they needed to do important philosophical spadework in order to make sense of how moral sentences could be true or false. In this article we'll look at the puzzle about moral truth as it is faced by early noncognitivists and by metaethical expressivists, the early noncognitivists' contemporary cousins. We'll look at what it would take for expressivists to 'earn the right' to talk about moral truths at all, and in particular at what it would take for them to earn the right to claim that moral truths behave in the ways that we should expect – including that meaningful moral sentences which lack presuppositions *are* true or false, and that classically valid arguments are truth-preserving. And finally, we'll look at whether the assumptions about truth which expressivists need in order to do these things have any independent plausibility or merit.

1 **noncognitivism and truth**

To understand what noncognitivism is, it is useful to start with Ayer's *emotivist* theory from chapter six of *Language, Truth, and Logic*. Ayer's was not the first noncognitivist theory, and not even the first noncognitivist theory in English-speaking philosophy¹; but it was a relatively early noncognitivist theory, and Ayer's explanation of its commitments is still fresh and illuminating, and is a good point of departure for understanding later noncognitivist views.

Ayer explained his theory by comparing the meanings of sentences involving 'wrong' to counterpart sentences which don't involve 'wrong'. He compared **1** to **2**, and **3** to **4**:

- 1** You acted wrongly in stealing that money.
- 2** You stole that money.

- 3** Stealing money is wrong.
- 4** Stealing money.

Each of these pairs, Ayer alleged, have the same *significance* – that feature of sentences which must pass the verifiability condition, according to Ayer's verificationist theory of meaning.

So the main feature of Ayer's view was that whatever kind of meaning 'wrong' has, when we add it to 'you stole that money' to get 'you acted wrongly in stealing that money', the significance is not affected. And similarly, when we add it to 'stealing money' to get 'stealing money is wrong', its significance is not affected, either. Intuitively, this is because neither of these sentences is about something more – wrongness – than their counterparts. They are just different – emotionally loaded – ways of saying the same thing. Ayer also suggests that we could perform the same function just as well with a special sort of exclamation points – instead of saying 'stealing money is wrong', we could instead just say, 'stealing money!!!' in a special tone of voice, or with especially thick exclamation points. The word 'wrong' doesn't affect *what* we say; it only affects *how* we say it.

Since significance is the aspect of meaning which is subject to Ayer's verificationism, Ayer is committed to holding that since sentence **2** can be verified, sentence **1** can be verified as well – under the very same conditions. Since to verify a sentence is presumably to verify that it is true, that means that Ayer's theory predicts that sentence **1** *can* be true – but only under the very same conditions that sentence **2** is true. In contrast to sentence **1**, however, sentence **3** would be true only under the same conditions that

¹ Axel Hägerström is usually credited with providing the first clear statement of emotivism, followed years later by Ogden and Richards [1923], which is usually cited as the first clear statement of emotivism in English-speaking philosophy. *Language, Truth, and Logic* was published in 1936.

sentence **4** would be true – but **4** isn't the kind of sentence to be true at all, for it fails to express a complete thought. So Ayer's view seems to predict that sentence **3** can't be true at all – and neither can it be false (since sentence **4** can't be false, either).

These are highly counterintuitive implications. It is counterintuitive that 'stealing money is wrong' cannot be true not only because we are ordinarily quite comfortable calling it true (if we agree with it – and calling it false if we disagree with it), but also because – as some of Ayer's earliest critics pointed out – arguments involving 'stealing money is wrong' as a premise or conclusion are sometimes valid, but an argument can be valid only if the truth of its premises guarantees the truth of its conclusion. If valid arguments need to be truth-preserving, then it is particularly problematic for it to turn out that 'stealing is wrong' cannot be true, because that means that it can't turn out to be the conclusion of a valid argument.

Even though Ayer's view makes room for the thesis that sentence **1** *can* be true, moreover, it seems to say the wrong things about *when* sentence **1** is true. This is so not only because it makes perfect sense for someone (think: Robin Hood) to think that **2** is true but that **1** is not true (if you are having trouble with this one, substitute ('giving to charity' or 'saving lives' for 'stealing money')). It is also so because since arguments are valid when the truth of their premises guarantees the truth of their conclusion, Ayer's view makes the wrong predictions about which arguments are valid. For example, it predicts the following argument to be valid:

- P1** You stole that money.
- P2** If you acted wrongly in stealing that money, then you ought to give it back.
- C** You ought to give it back.

But this argument is patently not valid (again, if you balk at this, substitute 'saving lives' for 'stealing money'). So for both kinds of sentences, Ayer's view makes highly counterintuitive predictions.

Now because it seemed to be such a clear consequence of Ayer's theory that sentences like 'stealing money is wrong' cannot be true, philosophers have often *defined* 'noncognitivism' to mean the view that moral sentences cannot be true or false. Clearly, in light of the obvious predictions of Ayer's view about sentence **1**, this formulation can't be right even about Ayer. But something roughly in its ballpark was for a long time taken to be definitive of a family of theories which, like Ayer's, held moral words like 'wrong' to make a different kind of contribution to the meaning of sentences than other kinds of words – more like the kind of contribution made by exclamation points or the addition of 'damn' than like the addition of 'square' or 'red' or 'ten' or 'electron'. The view, very roughly, is that adding moral words to a sentence

doesn't allow you to *say* anything different; what it allows you to *do*, is to *do* something different with what you say.

2 the deflationist move

More recently, however, philosophers who are sympathetic to Ayer's idea that moral words make a different kind of contribution to the meaning of sentences than do words like 'square' or 'electron' have observed that it seems in principle possible to hold this view while still saying that moral sentences can be true or false. After all, they have suggested, Ayer's basic idea wasn't that moral sentences can't be true or false – that was a consequence which he derived from an underlying picture about what the meaning of moral sentences is like. But maybe that consequence *doesn't* really follow. In that case, we could adopt Ayer's underlying view about what the meanings of moral sentences are like, but deny his view that moral sentences can't be true or false. These philosophers either rejected the label 'noncognitivism', which was associated with the definition, 'moral sentences can't be true or false', or sought to reclaim it, by arguing that it was better associated with the underlying ideas about meaning which had seemed to motivate the conclusions about truth.

The main idea of these philosophers was to adopt ideas from the *deflationist* theory of truth. Now the core ideas of deflationism can be and have been characterized in a number of closely related ways, but for our purposes a simple and not inaccurate way of understanding the basic idea, is that the word 'true' is used to agree, and the word 'false' is used to disagree – and that once you have a theory which can explain those two things, there is not really anything more to truth than that. Adopting this deflationist idea, the reclaimers argued that if you think that stealing money is wrong, then you agree with 'stealing money is wrong'. Since there is nothing more to calling it 'true' than agreeing with it, correct use of the word 'true' therefore permits – indeed, requires – calling it true. So well-brought-up philosophers like us, who *do* think that stealing money is wrong, should agree that it is true that stealing money is wrong. Consequently, we should agree that it can be true that stealing money is wrong. So we should admit that moral sentences can be true – and a similar argument shows that we should admit that moral sentences can be false.

According to these noncognitivist reclaimers of moral truth, their dispute with cognitivist theories of the meaning of moral sentences is less like the dispute between theists and atheists, and more like the dispute over Malebranchean *occasionalism*. According to Malebranchean occasionalism, not only does God exist, but He is the proximate cause of everything that happens. Malebranchean occasionalists aren't just theists; they are theists who claim that God does an awful lot of work. One way of disagreeing with

Malebranchean occasionalism is to be an atheist, and deny that God exists at all. But another way to disagree with Malebranchean occasionalism, is to believe in God, but simply hold that He does less work. The noncognitivist reclaimers of moral truth say that their dispute with cognitivist theories isn't about whether there are moral truths, but about how much *work* moral truth needs to do. Cognitivist theorists are like Malebranchean occasionalists in offering truth-conditional theories of the meaning of moral sentences, using truth to explain meaning. One way of disagreeing with cognitivism – as exemplified by some of the earliest noncognitivist theories – is like atheism; it is to deny that there even are any moral truths. But another, equally good, way of disagreeing with cognitivism, is to believe in moral truths, but simply to hold that they do less work. This is the path of the noncognitivist reclaimers of moral truth.

The idea that noncognitivists can say all of these things has now become commonplace in moral philosophy. But it leaves several challenges remaining. Here I'd like to catalogue three, and then explore in the remainder of this article how the main contemporary cousin of Ayer's noncognitivism, *expressivism*, addresses those three challenges.

The first challenge is that it is too quick to say that it follows from the basic deflationist idea that if you are willing to accept or assert a sentence, then you should be willing to accept or assert that it is true. This certainly not true, for example, of some of the sentences to which early noncognitivists like Ayer analogized moral sentences. For example, consider the sentence, 'Boo, stealing money!', or 'Stealing money!!!', with its specially thick exclamation marks. These are not the kind of sentence that can be true or false. Even if you are willing to assert them, you need not be willing to agree that they are true, because they are not the right *kind* of sentence to be true. But Ayer explicitly endorsed the view that 'stealing money is wrong' has exactly the same kind of meaning as 'Boo, stealing money!' or as 'Stealing money!!!' So there is a serious challenge to explaining why we should say that the former sentence can be true, but the latter can't. This challenge is I think quite a bit more difficult than some people have made it out to be; I won't have very much *new* to say about it in this article, but I will explain how it can be seen as playing a role in motivating expressivism.

The second challenge that I want to explore is that there is something unsatisfying about the initial explanation that we have received so far of why moral sentences can be true or false. The explanation given so far tries to show that consistency requires admitting that it is true that stealing money is wrong, if you think that stealing is wrong. So since we well-brought-up people think that stealing money *is* wrong, we are committed, on pain of inconsistency, to admitting that it is true that stealing money is wrong – and hence to admitting that there is a truth about whether stealing money is wrong. The reason why this is so far

unsatisfactory, is that it is consistent with all of this that since it is *not* true that stealing money is wrong, we should not think that stealing money is wrong. So far as this goes, someone who has no view about whether stealing money is wrong may perfectly well deny that there is or could be any truth of the matter about whether stealing money is wrong, and not be going wrong in any way. Another way of putting this worry, is that so far as the explanation we've seen goes, the claim that there is a truth about whether stealing money is wrong seems to be, at best, itself simply just another *moral* truth. It is something we are committed to since, being well-brought-up, we think that stealing money is wrong. The challenge is therefore to explain why even a moral agnostic would be mistaken to deny that there is a truth of the matter about whether stealing money is wrong. I think this is an easier challenge than the first, but I have never seen it raised or its answer spelled out clearly, so I will do that later in this article.

The third challenge I'm going to focus on, is the second of the two kinds of counterintuitive consequences of Ayer's view that sentences **1** and **3** had the same truth-values as **2** and **4**. The first counterintuitive consequence was that it didn't seem to match what we ordinarily say – and we've already seen how the deflationist can dismiss that consequence. If we will say that 'stealing money is wrong' is true just in case we agree that stealing money is wrong, and that 'you acted wrongly in stealing that money' is true just in case we agree that you acted wrongly in stealing that money, then this gets the intuitively right results about when we say these things. But that leaves open whether we deal with the second counterintuitive consequence isolated earlier – namely, that Ayer's view got the wrong results about which arguments are valid. The third challenge is to show not only that arguments with moral premises or conclusions are not *precluded* from being valid by lacking truth values or having the intuitively wrong truth values, but that a noncognitivist view can actually get the *right* results about which arguments are valid – in the sense of being truth-preserving. Simon Blackburn calls this the project of 'earning the right to truth', and I'll be illustrating in detail how noncognitivists aspire to do this, by walking through the way that it can be done by expressivism.

3 the first challenge and truth-aptness

Our first challenge is the challenge of drawing a line around *which* sentences we are supposed to admit are true, if we are willing to endorse them in the first place. The noncognitivist wants to say that 'stealing money is wrong' is among those to which this deflationist idea applies, but that 'Boo, stealing money!' is not. The mere fact that you are willing to go around saying, 'stealing money is wrong', requires you to be willing to assent to 'the sentence, 'stealing money is wrong', is true' as well, but the mere fact that you are

willing to go around saying 'Boo, stealing money!' does not require you to be willing to assent to 'the sentence, 'Boo, stealing money!', is true'. Since the latter sort of sentence *can't* be true, the challenge is to draw the line around the sort of sentences to which the deflationist idea applies in a way that includes 'stealing money is wrong' but does not include 'Boo, stealing money!'.

Another name for this challenge is the problem of articulating a satisfactory theory of *truth-aptness*. We may say that a sentence is *truth-apt* if it is the sort of sentence that, when you are willing to say it, you should be willing to allow that it is true. Now, deflationism about truth alone can't establish that moral sentences can be true or false – only deflationism together with the view that moral sentences are truth-apt can establish that those of us who accept some moral sentences, must admit that moral sentences can be true. Now every version of deflationism needs a theory of truth-aptness – a story about which sentences 'P' can be replaced with 'it is true that 'P' or with "'P' is true'. But noncognitivists face a special challenge in outlining a theory of truth-aptness, because a noncognitivist theory of truth-aptness must classify moral sentences like 'stealing is wrong' differently from sentences like 'Boo, stealing money!' and 'Dammit – not stealing money!', even though the central idea of noncognitivism, as we've seen, is that the former sentences have the very same kind of meaning as the latter sentences. The fact that noncognitivists think these sentences have the very same kind of meaning – perhaps even the *same* meaning – is the principal challenge to any theory of truth-aptness being able to classify moral sentences as truth-apt, but not the others.

Now, there is a very simple way in which some philosophers have believed that this trick could be accomplished. According to these philosophers, the distinction between truth-apt and non-truth-apt sentences is not a semantic distinction at all. It is simply a *syntactic* distinction. The test for whether a sentence 'P' is truth-apt or not, they have said, is simply whether 'it is true that P' is grammatical (in the sense of being syntactically well-formed) or not. Since on this view the distinction between truth-apt and non-truth-apt sentences is not a semantic one at all, but a purely syntactic one, it runs into no difficulty at all with the hypothesis that 'stealing money is wrong' and 'Boo, stealing money!' have *exactly* the same meaning, but the former is true and the latter is not.

This is because despite the fact that these two sentences are semantically identical, they are syntactically quite different. Indeed, the sentence 'it is true that stealing money is wrong' is grammatical, but the sentence 'it is true that Boo, stealing money!' is not. So this syntactic theory of truth-aptness not only tells us which sentences are truth-apt, and not only classifies this pair of sentences in the way that the noncognitivist hopes for, it does so in a way that is compatible with the very strongest sort of noncognitivism – with the view that these two sentences have exactly the same meaning.

Unfortunately, there are problems with the idea that truth-aptness is a merely syntactic property, rather than having any basis in semantics. This idea fares pretty well with how we actually classify existing sentences of English, sure enough, but it also predicts that all that we need to do in order to make a non-truth-apt sentence truth-apt is to change its grammar. Dreier [1996] tested this prediction by introducing a new predicate, 'hiyo'. According to Dreier, 'Bob is hiyo' is used in very much the same way as 'hey, Bob' is used – to accost Bob. But according to Dreier, 'hiyo' is grammatically an adjective, so 'Bob is hiyo' is an ordinary indicative sentence, and consequently 'it is true that Bob is hiyo' is syntactically well-formed. Consequently, by the syntactic criterion, 'Bob is hiyo' is truth-apt, and so someone who is willing to accost Bob should be willing to admit that 'Bob is hiyo' is true. Dreier calls this view about 'hiyo' *accostivism*.

It is far from clear, however, whether this should really be so. In the case of moral sentences, we certainly do ordinarily say that they can be true or false. And they are certainly indicative. But what is at issue, here, is precisely whether it is their indicative syntax alone which suffices to guarantee that they can be true or false. And it isn't clear whether 'Bob is hiyo' really should be said to be true or false, simply because we are willing or unwilling to accost Bob.

Dreier sharpens the worry by noting that sentences like 'if Bob is hiyo, then I'm out of here' and 'everyone who isn't hiyo is boring' are *also* grammatically well-formed, if 'hiyo' is an ordinary predicate – and indeed, are themselves truth-apt, by the syntactic criterion for truth-aptness. Moreover, since these sentences are formed truth-functionally on the basis of their parts (assuming for the sake of argument the material conditional treatment of 'if...then' for the former case), that tells us under what conditions these sentences *are* true. But as Dreier notes, even once we understand that 'Bob is hiyo' is used to accost Bob, so that we understand how to use *that* sentence, we *still* don't understand what sentences like 'everyone who isn't hiyo is boring' mean, or know how to use them. A theory which tells us that it is this easy to confer a meaning on these sentences makes too much out of syntax.

Dreier's discussion pushes the noncognitivist to explain the difference between 'stealing is wrong' and 'Boo, stealing!' on *semantic* grounds. The bare syntactic criterion on truth-aptness makes it too easy for sentences to turn out to be true or false, and consequently too easy to generate putatively meaningful complex sentences which are patently not meaningful at all. What noncognitivists need, in order to avoid the result that 'hiyo' is a perfectly intelligible predicate, is some kind of semantic constraint which, intuitively, goes hand-in-hand with indicative syntax in natural languages, but is violated when Dreier tries to simply stipulate indicative syntax for sentences involving 'hiyo'.

The best existing answer to Dreier's challenge that I know of comes from Allan Gibbard. Gibbard's idea is closely connected to my original characterization of deflationism as the idea that 'true' is

used to agree, and 'false' is used to disagree. The rough idea is that some sentences are not used to do things that it is possible to agree or disagree with – and that such sentences are not truth-apt. According to Gibbard, it is not possible to agree or disagree with an accosting, and so it is not possible to agree or disagree with 'Bob is hiyo'. Consequently, it does not make sense to say that 'Bob is hiyo' is true or false, even if you are willing or unwilling to accost Bob – and even though it is syntactically indicative.

In the abstract, it is obscure to talk about agreeing or disagreeing with what a sentence is 'used to do'. I used that expression in order to give a sense for Gibbard's answer to Dreier. But to fully understand Gibbard's answer to Dreier, we need to understand the *kind* of noncognitivist theory which Gibbard endorses, which is known as *expressivism*. In the next section I'll explain the basic ideas of expressivism, and then restate Gibbard's answer to Dreier in that framework. Then in the remainder of the article we'll go on to address the second and third challenges in the context of expressivism.

4 contemporary noncognitivism: expressivism

In my view, it is most promising and helpful to think of expressivism as a kind of *assertability conditional* semantic theory. On this interpretation, the main idea of expressivism is that just as the phonological rules of a language set *phonological* correctness conditions, and the syntactic rules of a language set *syntactic* correctness conditions (for example, 'colorless green ideas sleep furiously' meets these but 'example one this good a is' fails them, even if we can guess what it means), the semantic rules of a language set *semantic* correctness conditions. But on this view, semantic correctness conditions aren't *truth* conditions. To see why not, imagine that Todd followed the Drudge Report up until the end of October 2008, but then subsequently stopped paying attention to the US presidential election or its outcome, and consequently believes that John McCain was on track to become, and now is, the president of the United States. When asked who is the president of the United States, Todd volunteers, 'John McCain is president of the United States'. What Todd has said is not true, and so in asserting this he has made a mistake. But intuitively, Todd's mistake is not a linguistic one; it is a mistake about US electoral politics.

Examples like this one motivate the view that the semantic correctness condition of 'P' is not that P, but rather that the speaker thinks that P. It is because Todd thinks that John McCain is president of the United States that when he says so, he makes no semantic mistake. In general, on this view, a semantic theory needs to associate each sentence, 'P', of the language with a mental state: what it is to think that P. The semantic correctness condition of uttering 'P' is then that the speaker is in that mental state. We say that the state associated by the semantics with a sentence is the state that that sentence *expresses*.

What makes expressivism a way of developing noncognitivism, is the observation that within a semantic framework of this kind, it need not turn out that what it is to think that P is always to have an ordinary belief about the world. According to metaethical expressivism, when 'P' is an ordinary descriptive sentence like 'grass is green', what it is to think that P is to have an ordinary descriptive belief – in this case, the belief that grass is green. But when 'P' is a moral sentence like 'stealing money is wrong', what it is to think that P is not to have an ordinary belief about the wrongness of stealing money at all. Instead, it is to have a negative desire-like attitude toward stealing money – to *disapprove* of stealing money.

Disapproval of stealing money is only an attitude toward stealing money – and not an attitude toward wrongness. So it differs importantly from the belief that grass is green, which is an attitude not just toward grass, but toward its greenness. So metaethical expressivists agree with Ayer that moral words like 'wrong' make a different kind of contribution to the meanings of the sentences involving them than do ordinary descriptive words like 'green'. Rather than contributing to what accepting such sentences is having an attitude *about*, moral terms like 'wrong' affect which *kind* of attitude is involved in accepting sentences involving it.

In the terms of expressivism, Gibbard's response to Dreier is that truth-apt sentences must express states of mind with which it is possible to agree or disagree. It is possible to agree or disagree with ordinary descriptive beliefs, like the belief that grass is green. You disagree with it, for example, if you believe that grass is not green. In general, any two beliefs with inconsistent contents 'disagree' with one another, in the sense that anyone who had the first of these beliefs would disagree with anyone who had the second. So ordinary descriptive beliefs are uncontroversially the kind of state of mind with which it is possible to agree or disagree.

Following Stevenson, however, Gibbard holds that ordinary descriptive beliefs are not the only kinds of state of mind with which it is possible to agree or disagree. For example, it is also possible to disagree in intention. If the two of us are making plans about what to do tonight, I may intend that we go to the symphony, since I can't stand movies, and you may intend that we go to the cinema, since you can't stand classical music. If so, then intuitively, we disagree about what to do tonight. Gibbard holds that disapproval of stealing money is also an attitude that it is possible to disagree with. For example, if you disapprove of stealing money, and I disapprove of *not* stealing money, then we seem to be in some sort of disagreement. If this is right, then disapproval is the right sort of attitude to be expressed by truth-apt sentences. In contrast, it is not possible to disagree with a headache. So no sentence which expressed the state of having a headache could be truth-apt. And similarly, it is not possible to disagree with an accosting. So Dreier's stipulations about 'hiyo' do not suffice to make 'Bob is hiyo' truth-apt.

So now we've encountered the very basic idea of expressivism, and seen how it enables us to preserve some of the basic ideas underlying Ayer's emotivism, but in a way that enables a well-motivated answer to Dreier, and consequently a good explanation of how 'stealing money is wrong' could be truth-apt, even though 'Boo, stealing money!' and 'Bob is hiyo' are not. In section 5, I'll explain the main problem facing noncognitivism – including expressivism – which is often known as the *Frege-Geach* problem, and what it has to do with truth. That will set us up to better understand what expressivist attempts to explain validity are like and what an expressivist account of truth would look like, in sections 6 and 7. Finally, in section 8 we'll reap the payoff of this extended discussion of expressivism, by seeing how expressivists can hope to address my second and third challenges. And in section 9 we'll examine the undischarged assumptions we needed to make on behalf of expressivists along the way.

5 the frege-geach problem

The most famous and pressing problem facing any version of noncognitivism – expressivism included – is that of providing an adequate compositional semantics: of accounting for the meanings of complex sentences, and doing so on the basis of the meanings of their parts. To get an initial sense for this problem, recall Ayer's view that 'wrong' does not add anything to the significance of sentences in which it figures, and look at a couple of sentences which are even less congenial to this proposal than the examples that he discusses:

- 5** Everything wrong is forbidden in the Bible.
- 6** Everything is forbidden in the Bible.

- 7** Stealing money is wrong or my parents lied to me.
- 8** Stealing money or my parents lied to me.

Obviously here we have a problem; sentence **5** is not merely a different way of saying sentence **6** – on the contrary, most people who would assent to **5** would deny **6**. Even worse, sentence **8** isn't even well-formed; you can't take the disjunction of a gerundival phrase and a sentence. Certainly if Ayer's view is committed to thinking that you can, then he has his work cut out for him in explaining to us what it means.

The problem is that it is all well and good for Ayer to tell us what simple sentences involving 'wrong' mean, by comparing 'stealing money is wrong' to 'stealing money!!!'. But an adequate theory of meaning must not only tell us the meanings of simple sentences; it must also provide us with a way of

determining the meanings of complex sentences. But as the examples just considered illustrate, Ayer did not get very far in considering the implications of his view for complex sentences, and as the example of sentence 7 illustrates, the ordinary compositional rules which work for other sorts of semantic theory do not look like they are capable of doing the sort of work which Ayer's view requires of them.

In the early 1960's, Peter Geach pressed this problem against noncognitivism as he interpreted it.² What Geach did, was to collect pieces of evidence that moral sentences have the same meaning, even when they appear as part of complex sentences. For example, compare 'stealing money is wrong' to 'is it the case that stealing money is wrong?' The former of these sentences is an *answer* to the latter, and that is so precisely because the words 'stealing money is wrong' means the same thing in both sentences – if any of those words were ambiguous across the two sentences, like the ambiguity in 'bank', then the former sentence would not be the answer to the latter. Another example Geach considered was the connection between 'stealing money is wrong' and 'it is not the case that stealing money is wrong'. Again, Geach argued, 'stealing money is wrong' means the same thing as it appears in both sentences – and the evidence for that is that the two sentences are *inconsistent*.

The most famous example used by Geach was intended to show precisely the same thing, about a different pair of sentences. He considered sentences like 'stealing money is wrong', 'if stealing money is wrong, then embezzlement is wrong', and 'embezzlement is wrong'. Geach noted that the first two sentences provide a valid argument for the third, and again contended that this is because 'stealing money is wrong' has the same meaning in both places in which it appears, and that 'embezzlement is wrong' has the same meaning in both places in which it appears.

Geach made a big deal out of his arguments that the words 'stealing money is wrong' have the very same meaning when they appear in complex sentences as they do when they appear unembedded, because he believed that this was inconsistent with noncognitivism. As Geach understood noncognitivism, what gives moral sentences the meaning that they have, is what they are used to *do*. For example, Ayer seemed to believe that what gives 'wrong' its meaning, is the emotion it is used to express; similarly Stevenson seemed to believe that what gives 'wrong' its meaning, is the emotion it is used to encourage one's audience to have; and Hare seemed to believe that what gives 'wrong' its meaning is that it is used to condemn. But, Geach noted, the question, 'is stealing money wrong?' is not used to express the same negative emotion as 'stealing money is wrong', nor is it used to encourage one's audience to have that same emotion, nor is it used to condemn stealing. So, Hare, concluded, noncognitivists are committed to denying an obvious and

² John Searle [1962] independently pressed a very similar objection, but for Searle the dialectic was slightly more complicated; see Schroeder [2008c] for discussion.

important truth: namely, that 'wrong' means the same thing when it appears in questions as when it appears in indicative sentences. And he made similar points for negations and conditionals.

The consensus response to the problem as posed by Geach, is that moral sentences like 'stealing money is wrong' don't have to be used to do the very same thing, in complex sentences, as they are used to do when unembedded, in order to count as having the same meaning. All that has to happen, is that the meaning of the complex sentences in which they figure is a *function* of the meanings of the constituent sentences. This is the answer which Hare gave to Geach (and to Searle, who offered a similar argument against noncognitivism) in 1970, and it has informed research on the Frege-Geach problem ever since.

In light of Hare's response to Geach, we can think of the problem facing noncognitivism (including expressivism) as being that of providing compositional rules which determine the meanings of questions, negations, and conditionals (to say nothing of other sorts of complex sentences) as a function of the meanings of their parts – and all within a noncognitivist semantic framework. Within an expressivist theory, what this means is that an account of the meaning of 'not' must provide a function from mental states to mental states, which associates the mental state expressed by any sentence 'P' with the mental state expressed by its negation, ' \sim p'. Similarly, an expressivist account of the meaning of 'if...then' must provide a function from pairs of mental states to mental states, which associates the mental states expressed by any pair of sentences, 'P' and 'Q', with the mental state expressed by the conditional formed from them, 'P \rightarrow Q'. And so on for other complex-sentence-forming constructions.

The central constraint on this enterprise derives from the fact that not just any functions will do; these functions must assign the *right* mental state to the complex sentences, in a way that allows us to predict and explain the semantic properties of complex sentences. And Geach's argument provides us with an initial list of just what those semantic properties are: yes-no questions should turn out to be answered by their indicative counterparts, negations should turn out to be inconsistent with the sentences they negate, and conditionals should turn out to license *modus ponens*. No compositional semantic theory will be adequate unless it can explain these basic things about each of these constructions.

It is important to understand that the Frege-Geach Problem is not fundamentally a problem about truth. It is a fully general problem about whether a non-standard semantic theory like expressivism can account for the compositional features of natural languages. But truth does come into the picture, because some of the important semantic properties of sentences formed by the connectives of propositional logic (for example) *do* seem to have to do with truth. For example, the most important semantic property of 'not' seems to be that negations are inconsistent with the sentences that they negate – i.e., that 'stealing

money is not wrong' is inconsistent with 'stealing money is wrong'. But inconsistency would ordinarily be defined in terms of truth – two sentences are inconsistent just in case they can't both be true. Similarly, the most important semantic property of conditionals is usually taken to be that they validate *modus ponens*. But validity would ordinarily be defined in terms of truth – an argument is valid just in case the truth of its premises guarantees the truth of its conclusion. And so these observations get us back to the problem: just how are expressivists going to make sense of moral truth?

The usual expressivist strategy for explaining these things is to divide the semantic properties of sentences up into two classes, a primary class, and a secondary class. The primary class of semantic properties of sentences includes properties about the rational relationships between *accepting* the sentences (more on this in a moment); the secondary class of semantic properties of sentences includes all of their other semantic properties, including facts about which sentences they are inconsistent with (in the sense that both can't be true) and which valid arguments they figure in (in the sense of which preserve truth). Then expressivists divide their theory up into two stages. At the first stage, they propose compositional theories of which mental states are expressed by complex sentences as a function of the mental states expressed by their parts, and try to use those functions in order to explain the primary class of semantic properties of those sentences. At the second stage, they try to *use* the primary semantic properties of sentences in order to predict and explain their secondary semantic properties – including facts about inconsistency and validity.

Much of the attention that has been paid to expressivist theories and in particular to expressivist solutions to the Frege-Geach Problem has focused on the first of these two stages – on which mental states the theory says are expressed by complex sentences. In what follows, I will instead focus on the second of these two stages. I will be offering an approach to understanding what happens at this second stage of expressivist theorizing – an approach that I call *commitment theory*. I believe that commitment theory, as I will be developing it here, is neutral between all existing expressivist theories. Studying commitment theory illustrates two things: first, how much expressivists can hope to accomplish – including providing the answers to our second and third challenges – if only they can manage to discharge the first stage of their theory. And second, it shows us exactly what intermediate goal the first stage of expressivist theorizing needs to meet, in order to be able to have these further payoffs.

We can therefore think of commitment theory as trying to reduce to a very manageable size the number of things which expressivists need to be able to explain at the first stage of their theory, and showing how much more can be explained on the basis of those resources. In sections 6 and 7 I'll explain

how commitment theory works, and then in section 8 we'll apply it in order to answer our second and third challenges about moral truth.

6 commitment theory – the basics

Commitment theory assumes that since expressivism assigns each sentence, 'P' to a mental state – to what it is to think that P – it should be in principle possible to use general principles from the philosophy of mind in order to explain the rational relationships between these mental states. So the idea is that a good place for expressivists to start, in explaining the semantic properties of a sentence like 'stealing money is not wrong', is to start by explaining the rational relationship between thinking that stealing money is not wrong, and having other sorts of thought. In particular, a good place to start would be by explaining why it is *rationally inconsistent* to think these two things at the same time.

Note that the sense in which it is rationally inconsistent to think that stealing money is wrong and also think that stealing money is not wrong, is not just that 'stealing money is wrong' and 'stealing money is not wrong' are inconsistent. Even though these are inconsistent, there is no rational inconsistency in supposing that stealing money is wrong and also supposing that stealing money is not wrong, or in wondering whether stealing money is wrong and also wondering whether stealing money is not wrong. So there is more to *rational inconsistency* in the sense at stake, here, than merely a matter of having thoughts which can't both be true. Gibbard [2003] calls the kind of rational inconsistency that we are interested in *disagreement*; the idea is that when you think that stealing money is wrong and someone else thinks that it is not wrong, you disagree with her; similarly to think both at the same time would be to disagree with yourself.

The fact that it is rationally inconsistent to think that stealing money is wrong and to also think that stealing money is not wrong is just a special case of a much more general fact. Whenever an argument is classically valid, it is rationally inconsistent to accept all of its premises and deny its conclusion. The original observation is a special case of this more general principle, because the single-premise argument whose premise and conclusion are both 'stealing money is wrong' is classically valid. Let us say that an argument has the *inconsistency property* just in case it is rationally inconsistent to accept its premises and deny its conclusion. One thing that expressivists ought to hope to be able to explain, therefore, is the fact that all classically valid arguments have the inconsistency property.

A related but different property that an argument might have is the property of being *inference-licensing*. Let us say that an argument *licenses inference* just in case someone who accepts its premises is *committed*

to accepting its conclusion. Since all theorems of classical logic are the conclusions of classically valid arguments from vacuous premises, it follows that if all classically valid arguments license inference, then everyone is committed to *accepting* all classical theorems, no matter what else they accept. Similarly, since all classically valid arguments have the inconsistency property, it is rationally inconsistent for anyone to deny any classical theorem, no matter what else they accept.

So long as it is rationally inconsistent to accept and deny one and the same sentence, the former of these two claims entails the latter. But the converse entailment is more controversial. Some philosophers believe that it sometimes makes sense to *reject* a sentence, neither accepting nor denying it, but not merely withholding, or simply waiting to make up one's mind, either. If it makes sense to reject 'P', some of these philosophers believe, then it also makes sense to reject ' $P \vee \sim P$ '. On this view, it is always inconsistent to *deny* a classical theorem, but we are not always committed to *accepting* them. So on this view, not all classically valid arguments license inference – in particular, arguments which rely on steps with vacuous premises do not. On this picture, the inferences which are licensed are those whose conclusions are 'relevant' to their premises. So the inference-licensing property can come apart, at least in principle, from the inconsistency property.

Rather than evaluate whether there really is such an attitude as rejection, or whether it is ever rational to reject a sentence, I am going to leave this question open, by working with a framework that leaves the possibility of rejection open, but which also tells us what happens if there is no such attitude, or if it is never rational. I will assume, therefore, that there are *at most* three possible attitudes to have toward a sentence, 'P' – acceptance, rejection, and denial. I will also assume that denying 'P' is the same as accepting ' $\sim P$ ', and (this is the 'at most' part) that these are the *only* positive attitudes one can have toward 'P', in the sense that someone who does not have one of these attitudes toward P is undecided whether P. And finally, I will assume that each of these three attitudes is rationally inconsistent with either of the other two attitudes toward one and the same sentence. These are the main assumptions of commitment theory.

7 commitment theory – the substance

With those assumptions in hand, we can exploit the notion of a *commitment table*, in order to map out which of those three attitudes a thinker is committed to having toward a sentence, on the basis of the attitudes that she has toward other sentences. For example, compare the following table:

P	$\sim P$
A	D
R	R
D	A

This is the commitment table for ' \sim ' ('not'). It tells us that someone who accepts 'P' is committed to denying ' $\sim P$ ' (that is, to accepting ' $\sim\sim P$ '), that someone who rejects 'P' is committed to rejecting ' $\sim P$ ', and that someone who denies 'P' is committed to accepting ' $\sim P$ ' (in fact, by assumption, these two states are identical).

Don't be misled into thinking that this is a truth table; it is not, though it does look like one. The table tells us nothing about the semantic status of either 'P' or ' $\sim P$ ' – it only tells us which combinations of views about 'P' and ' $\sim P$ ' are rationally consistent, and which are not. A pair of views is rationally consistent if it corresponds to a row on the table; otherwise not. (In deriving commitment tables like this one, we can either take as primitive the notion of commitment, and explain rational inconsistency as being committed to having different attitudes to the same sentence, or we can start with a notion of rational inconsistency, and explain commitment, by saying that states $S_1 \dots S_n$ commit to attitude A to 'P' just in case of the three attitudes toward 'P', A is the only one that is not rationally inconsistent with the combination of $S_1 \dots S_n$.)

Similarly, the commitment table for '&' would be:

P	Q	P&Q
A	A	A
A	R	R
A	D	D
R	A	R
R	R	R
R	D	D
D	A	D
D	R	D
D	D	D

Together the commitment tables for ' \sim ' and '&' suffice to derive commitment tables for every sentence of propositional logic. Moreover, facts about these two tables suffice to establish that it is rationally inconsistent to deny any theorem of classical propositional logic. This is because these commitment tables are structurally identical to what are known as the strong Kleene 3-valued truth tables, and so it is a well-known fact that the commitment tables for all and only classical theorems never contain a 'D' – which

means that there are no consistent states of mind which involve denying them.³ (If we assume that rejection is not really a rational possibility, then this turns into the result that everyone is committed to *accepting* every classical theorem.)

Moreover, it is a fact about classical logic that whenever $P_1 \dots P_n; Q$ is a classically valid argument, $(P_1 \& \dots \& P_n) \supset Q$ is a classical theorem – which means that it is inconsistent to deny. But it is straightforward to prove from our commitment tables that someone who accepts all of $P_1 \dots P_n$ as well as $\sim Q$ is committed to denying $(P_1 \& \dots \& P_n) \supset Q$, given the appropriate definition of ' \supset ' in terms of '&' and ' \sim '. That is, they are committed to denying something that it is rationally inconsistent to deny, if the argument from $P_1 \dots P_n$ to Q is classically valid. So this suffices to establish that classically valid arguments all have the inconsistency property – just on the basis of our two commitment tables.

Similarly, our two tables suffice to establish that all *relevance*-valid arguments have the inference-licensing property, and that if rejection is not rationally possible, then all classically valid arguments have the inference-licensing property. This is because to say that an argument is inference-licensing is to say that accepting its premises commits you to accepting its conclusion – which is to say that the commitment table for the conclusion should have only 'A's, wherever there are 'A's for all of the premises. But it is a well-known fact about the strong Kleene truth tables that all 'relevance-valid' arguments have this feature.⁴ The following table illustrates the case for *modus ponens* (observe that every row with an 'A' for both 'P' and 'P \supset Q' also receives an 'A' for 'Q'):

P	Q	P \supset Q
A	A	A
A	R	R
A	D	D
R	A	A
R	R	R
R	D	R
D	A	A
D	R	A
D	D	A

³ For the proof, see, for example, Avron [1991]. This result is equivalent to the result that the logic LP has all and only the theorems of classical logic. This is because LP results from the strong Kleene truth tables under the treatment of both 'A' and 'R' as 'designated'.

⁴ For our purposes here, we may understand an inference to be relevance-valid only if the conclusion shares some propositional variable with the premises. The basic idea is that the only classically valid arguments for which our commitment tables don't assign an 'A' to the conclusion whenever they assign 'A' to all of the premises, are ones which rely on classical theorems involving sentences which don't appear in the premises – for example, as in the argument from 'P' to ' $\sim(Q \& \sim Q)$ '. We can think of ' $\sim(Q \& \sim Q)$ ' as an auxiliary premise which, though undeniable, need not be accepted – and hence someone who rejects it need not be committed to ' $\sim(Q \& \sim Q)$ ', even though she is committed to 'P' and the argument from 'P' to ' $\sim(Q \& \sim Q)$ ' is classically valid.

Similarly, if we assume that rejection is not rationally possible, then we can simply delete every row containing an 'R', and our commitment tables turn out to be structurally identical to the classical truth tables, with 'A' substituted for 'T' and 'D' substituted for 'F'. Obviously, every classically valid argument preserves 'T' – has only 'T's on its truth-table, wherever 'T' is assigned to all of its premises. So similarly, every classically valid argument preserves 'A' – has only 'A's on its commitment-table, wherever 'A' is assigned to all of its premises. That is, accepting its premises commits you to accepting its conclusion – it licenses inference.

It is a consequence of these observations, that these two commitment tables reduce the challenge facing expressivist accounts of the logical sentential connectives in the *first* stage of expressivist theorizing to the challenge of predicting and explaining these two commitment tables. The idea is to provide a recipe for what it is to accept ' \sim P' on the basis of what it is to accept 'P' which predicts and explains why accepting 'P' commits to rejecting ' \sim P' – that is, to accepting ' $\sim\sim$ P', and conversely. And (if you go in for rejection) provide a recipe for what it is to reject 'P' on the basis of what it is to accept 'P' which explains why rejecting 'P' commits to rejecting ' \sim P' and conversely, and is rationally inconsistent both with accepting 'P' and with accepting ' \sim P'. Finally, provide a recipe for what it is to accept 'P&Q' on the basis of what it is to accept 'P' and to accept 'Q' which, together with the other recipes, explains why accepting both 'P' and 'Q' commits to accepting 'P&Q' and conversely, and so on for each of the other properties of the commitment table for '&'. The idea is, so long as your expressivist account can explain the properties that are captured in the commitment table, then we can use those results (as we have here) in order to predict and explain at least one important property of all classically valid arguments. (The next step – for section 8 – will be to appeal to this property of classically valid arguments, in explaining why they are truth-preserving.)⁵

So while none of this, I think, tells us what the first stage of an expressivist semantic theory should look like – none of it tells us what it is to accept complex sentences involving ' \sim ' or '&' – it does set us a reasonable intermediate goal that seems like at least the *kind* of thing that an expressivist account ought to be able to aspire to explain. The idea is, we start with rational relationships between mental states, and then we try to explain other things from there. It is fruitful to start looking for how an expressivist account of truth would work, by looking in the same place: rather than starting from the beginning, by asking 'what is it to accept sentences involving 'true'?', the idea is to start mid-way, by trying to characterize some criteria

⁵ For an illustration of the first stage of an expressivist theory which does exactly this, see chapters 4-8 of *Being For*, especially chapter 8, section 1.

that an adequate answer to the former question might be expected to meet. That, in turn, will help to reduce the task faced by the former question and help us to better understand the constraints that it faces.

So we can think of an expressivist approach to truth in this same framework. In order to avoid issues about whether expressivists can quantify over propositions – including over moral propositions – I'll focus here on a sentential truth predicate. Now obviously, the commitments of someone who thinks a sentence S is true, depend on what she thinks S means. So the commitment table for 'S is true' will need to be sensitive to what you think S means. The relevant part of the table looks like this:

P	S means that P	S is true
A	A	A
R	A	R
D	A	D
A	R	
R	R	
D	R	
A	D	
R	D	
D	D	

I'm leaving the bottom part of the column for 'S is true' shaded, largely because it won't matter for our purposes how it is filled in – though it is worth noting that things are a little bit more complicated here, since the section where 'S means that P' receives a 'D' column will not be a function of the thinker's attitudes with respect to 'P' and to 'S means that P'.

This much of the table, however, is sufficient to establish that when someone accepts 'S means that P', she is committed to having the same attitude toward 'S is true' and toward 'P' – whatever that attitude is. Moreover, since her commitment toward complex sentences formed by the connectives of propositional logic is commitment-functional (as spelled out by the commitment tables for '~' and '&', that means that 'P' and 'S is true' can be substituted everywhere in sentences formed by such connectives, and preserve commitment (for someone who accepts 'S means that P'). This is a deeply important result. It means that 'P' and 'S is true' are completely intersubstitutable – that any inference substituting one for the other licenses inference, so long as 'S means that P' is also accepted.

8 applications: the second and third challenges disposed of

We can use this result about the intersubstitutability of 'P' and 'S is true' in order to derive many important observations about truth. For example, we can use it to show that for someone who thinks that 'stealing is wrong' means that stealing is wrong and 'stealing is not wrong' means that stealing is not wrong, "stealing is wrong" is true or 'stealing is not wrong' is true' will have the same commitment status as 'stealing is wrong or stealing is not wrong' – that is, it will be rationally inconsistent to deny it, and everyone will be committed to it, if rejection is not a rational possibility. So as long as someone admits that 'stealing is wrong' means that stealing is wrong, on this view she cannot rationally deny that there is a truth of the matter about whether stealing is wrong – even if she herself has no opinion about whether stealing is wrong or not.

This means that on this view, the claim that there is a truth of the matter about whether stealing is wrong is *not* simply a moral truth. It is not simply something that people will say, who either think that stealing is wrong, or who think that stealing is not wrong. It is something that *no one* who thinks that 'stealing is wrong' means that stealing is wrong and that 'stealing is not wrong' means that stealing is not wrong can consistently deny. The claim that there is a truth of the matter about whether stealing is wrong therefore has a special status due to the meaning of 'true' – not just to the fact that we well-brought-up folks in fact think that stealing is wrong. So that is an answer to our second challenge

Similar arguments can be used to explain why classically valid arguments are truth-preserving, answering our third challenge and completing the response to the Frege-Geach Problem by explaining how conditionals validate *modus ponens*. We already showed, above, that classically valid arguments have the inconsistency property, and in particular, that when $P_1 \dots P_n; Q$ is a classically valid argument, it is rationally inconsistent to deny $(P_1 \& \dots \& P_n) \supset Q$. But that means that for someone who thinks that 'P₁' means that P₁, ... 'P_n' means that P_n, and 'Q' means that Q, '((P₁ is true) & ... & (P_n is true)) \supset (Q is true)' will also be rationally undeniable.

So now we make the transcendental turn: since it is undeniable that if the premises of a classically valid argument are true, then the conclusion is true as well, let's not deny it. Moreover, since rejecting it commits us to rejecting at least one of its atoms, let's not reject it, either. Consequently, let us accept it. So since we accept it, let us say it: if the premises are true, then the conclusion is true, as well.

This is how expressivists can 'earn the right' to truth. The first step is to provide an expressivist account of truth, which tells us which mental states are expressed by sentences involving 'true' – that is, what it is to accept such sentences. We haven't explored that step, here – instead we have formulated an

intermediate stage, where we offered conditions of adequacy for what such an account should be able to establish, in the form of commitment tables for '¬', '&', and 'true'. The second step is to show how, given the first step, 'S is true' is substitutable everywhere for 'P', in the commitments of anyone who thinks that S means that P. The third step is to use that fact in order to show that the claim that if the premises of an argument are true, then the conclusion is true as well, has the same status as the theorems of classical logic. And finally, the fourth step is, since the claim that such arguments preserve truth has the same status as the truths of classical logic, to make the transcendental turn, and accept that claim.

9 what remains for expressivist truth?

Commitment theory doesn't tell us how expressivists can earn the right to truth; it only tells us how expressivists can *hope* to earn the right to truth. To make good on this hope, they need to make good on the first stage of an expressivist semantic theory, and tell us what mental states are expressed by complex sentences like '¬P', 'P&Q', and 'S is true', and to use their answers to those questions, along with background assumptions in the philosophy of mind, in order to predict and explain the commitment tables for '¬', '&' and 'true'. So commitment theory doesn't solve the problems for expressivism – either the Frege-Geach Problem or the problem about moral truth. But commitment theory does refine our understanding of the problems facing expressivism, and give us a clearer sense of how, by accomplishing some intermediate goals, an expressivist theory can eventually hope to be able to explain why it is not just a 'moral truth' that there is a truth of the matter about whether stealing money is wrong, and why even moral arguments – and the right moral arguments, at that – can be and are truth-preserving.⁶

It is important to appreciate, however, that in order to do what expressivists need of them, expressivist accounts of truth like the one sketched here in outline cannot just be something to be said about ethics. They have to be understood as quite general theories about the nature of truth – or more precisely, about the meaning of 'true' – *in general*. If the expressivist theory about the meaning of the word 'true' does not stand up to the standards requisite for evaluating any other theory of truth, then truth really does create a problem for expressivism, after all.

Fortunately for expressivism, there are at least some good things to be said in favor of the kind of account of truth sketched here. The most general of these, is that expressivism is a way of implementing a more general *deflationism* about truth. Recall that expressivists hold that 'wrong' does not contribute to what 'stealing money is wrong' is *about*, or to the *content* of the thought that stealing money is wrong, but rather

⁶ See Schroeder [2008b], chapter 11, and especially Schroeder [forthcoming] for a treatment of how expressivists might seek to make good on the first stage of this kind of theory, and explain the commitment tables for '¬', '&', and 'true'.

contributes to the *kind* of thought the thought that stealing is wrong is. Similarly, an expressivist account of 'true' holds that 'true' does not contribute to what 'S is true' is about, or to the content of the thought that S is true, but rather contributes to the kind of thought the thought that S is true is. This corresponds to the idea that just as wrongness isn't really a feature of the world, but just a feature of how we think about it, truth isn't really a feature of the world, either, but just a feature of how we think about it.

Among the many general reasons – applicable well outside of the study of ethics – for thinking that this might be the case about truth, is the famous paradox of the liar. The liar is a sentence which says of itself that it is not true. If truth is a feature of the world, then it would seem that there should be an answer to whether it is a feature that is possessed by the liar, or not. But neither answer seems to be satisfactory – indeed, a seemingly undeniable principle seems to predict that both answers are contradictory. The seemingly undeniable principle which yields this prediction is as follows:

T-schema If S means that P, then S is true just in case P.

The instance of this schema which leads to our problem, is therefore:

T-liar If liar means that liar is not true, then liar is true just in case liar is not true.

Many philosophers have noted not only that T-schema seems very hard to deny, but two other important things. (1) First, that even if it is not unrestrictedly true, all of its instances exert *pull*, in the sense that they *seem* to us like they have to be true.⁷ (2) And second, that we apparently have a *need* for the unrestricted T-schema, since we reason in ways that appear not to make sense if not all instances of the T-schema are true.⁸ This reasoning leads some of those philosophers to believe that every instance of the T-schema *is* true. And that, in turn, leads them to conclude that, since liar does mean that liar is not true, liar is true just in case liar is not true. At worst, following classical logic, this leads some philosophers to conclude that liar is true *and* not true, and at best, some philosophers give up on classical logic, simply in order to stop with a conclusion of the form 'P iff ~P', rather than getting to an outright conclusion, 'P and ~P'.

Now, it seems like it should be a relatively obvious observation that since even the problematic instances of the T-schema like T-liar seem undeniable, but affirming them leads to paradox, perhaps the right conclusion is that every instance of the T-schema *is* undeniable – though not every instance need be accepted. Perhaps the appropriate attitude toward T-liar is to *reject* it, rather than to deny it. The

⁷ See especially Eklund [2002].

⁸ See especially Gupta [2005] and chapter 13 of Field [2008].

expressivist account of truth sketched in the last two sections in terms of commitment theory makes good on this idea. It follows from that theory that every instance of T-schema *is* undeniable. For to deny an instance of T-schema, one would need to accept the antecedent and deny the consequent. That is, she must accept 'S means that P' and deny 'S is true if and only if P'. But the commitment-table for 'true' shows us that someone who accepts 'S means that P' has the same commitment for 'S is true' as for 'P'. So since that is so, she cannot consistently deny 'S is true if and only if P'. She can either accept it (if she either accepts or rejects both 'P' and 'S is true') or reject it (if she rejects both 'P' and 'S is true'), but if she denies it, then she is being rationally inconsistent.

So far, this falls far short of a vindication of the expressivist approach to truth; it is more like a reason to wonder whether the expressivist approach might turn out to be defensible on its own merits, after all. And even if the approach has some potential virtues, there is still the first stage of the expressivist theory to be carried out. But promise of being defensible on its own merits is what expressivists *need* of their account of truth. For the noncognitivist project of reclaiming moral truth is not just about moral truth; it is about the whole of truth.⁹

references

- Avron, Arnon [1991]. 'Natural 3-Valued Logics – Characterization and Proof Theory.' *The Journal of Symbolic Logic* 56(1): 276-294.
- Ayer, A.J. [1936]. *Language, Truth, and Logic*. New York: Dover.
- Blackburn, Simon [1984]. *Spreading the Word*. Oxford: Oxford University Press.
- _____ [1993]. *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- _____ [1998]. *Ruling Passions*. Oxford: Oxford University Press.
- Boghossian, Paul [1990]. 'The Status of Content.' *Philosophical Review* 99(2): 157-184.
- Dreier, James [1996]. 'Expressivist Embeddings and Minimalist Truth.' *Philosophical Studies* 83(1): 29-51.
- _____ [2006]. 'Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution.' *Oxford Studies in Metaethics*, volume 1. Oxford: Oxford University Press, 217-233.
- Eklund, Matti [2002]. 'Inconsistent Languages.' *Philosophy and Phenomenological Research* 64: 251-275.
- Field, Hartry [2008]. *Saving Truth from Paradox*. Oxford: Oxford University Press.
- Geach, P.T. [1958]. 'Imperative and Deontic Logic.' *Analysis* 18: 49-56.
- _____ [1960]. 'Ascriptivism.' *Philosophical Review* 69: 221-225.

⁹ Special thanks to Michael Glanzburg, Alexis Burgess, Matti Eklund, Barry Lam, Mike McGlone, Jake Ross, and Scott Soames.

- _____ [1965]. 'Assertion.' *Philosophical Review* 74: 449-465.
- Gibbard, Allen [1990]. *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.
- _____ [2003]. *Thinking How to Live*. Cambridge: Harvard University Press.
- Gupta, Anil [2005]. 'Do the Paradoxes Pose a Special Problem for Deflationism?' In J.C. Beall and Bradley Armour-Garb, eds., *Deflationism and Paradox*. Oxford: Oxford University Press, 133-147.
- Hale, Bob [1993]. 'Can There Be a Logic of Attitudes?' In Haldane and Wright, eds., *Reality, Representation, and Projection*. New York: Oxford University Press.
- Hare, R.M. [1970]. 'Meaning and Speech Acts.' *The Philosophical Review* 79(1): 3-24.
- Horgan, Terry, and Mark Timmons [2006]. 'Cognitivist Expressivism.' In Horgan and Timmons, eds., *Metaethics after Moore*. Oxford: Oxford University Press, 255-298.
- Horwich, Paul [1990]. *Truth*. Oxford: Oxford University Press.
- _____ [1993]. 'Gibbard's Theory of Norms.' *Philosophy and Public Affairs* 22: 67-78.
- Ogden, C.K., and I.A. Richards [1923]. *The Meaning of Meaning*. New York: Harcourt Brace.
- Price, Huw [1994]. 'Semantic Deflationism and the Frege Point.' In S. L. Tsohatzidis, ed., *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*. London: Routledge.
- Schroeder, Mark [2008a]. 'Expression for Expressivists.' *Philosophy and Phenomenological Research*
- _____ [2008b]. *Being For: Evaluating the Semantic Program of Expressivism*. Oxford University Press, May 2008.
- _____ [2008c]. 'What is the Frege-Geach Problem?' *Philosophy Compass* 3/4: 703-720.
- _____ [2009]. *Noncognitivism in Ethics*. New York: Routledge.
- _____ [forthcoming]. 'Expressivist Truth and the Liar.' Forthcoming in *New Waves in Truth*, edited by Nikolaj Jang Pedersen and Cory Wright.
- Searle, John [1962]. 'Meaning and Speech Acts.' *Philosophical Review* 71:423-432.
- Smith, Michael [1994]. 'Why Expressivists About Value Should Love Minimalism About Truth.' *Analysis* 54(1): 1-12.
- Stevenson, C.L. [1937]. 'The Emotive Meaning of Ethical Terms.' Reprinted in Stevenson [1963], 10-31.
- _____ [1944]. *Ethics and Language*. Oxford: Oxford University Press.
- _____ [1963]. *Facts and Values*. Westport, CT: Greenwood Press.
- Stoljar, Daniel [1993]. 'Emotivism and Truth Conditions.' *Philosophical Studies* 70: 81-101.
- van Roojen, Mark [1996]. 'Expressivism and Irrationality.' *The Philosophical Review* 105(3): 311-335.
- Wright, Crispin [1992]. *Truth and Objectivity*. Cambridge, MA: Harvard University Press.