

VII Problems for Expressivism

1. Can we really give up on realism?

A fundamental philosophical question for expressivism is whether it really gives us a plausible account of what we ordinarily take ethics to be, or whether what it gives us is something that falls short. To try to get a grip on this, take an issue that you really morally care about: whether it be issues of racism, or sexism, our treatment of animals or the environment, the status of the unborn child, our responsibilities to the rest of the world; everyone has something that they are deeply morally committed to. And then think whether there is really nothing more to the

There are at least two issues here. One is disagreement. Suppose you are arguing with someone who disagrees with you on one of these issues. Stevenson points out that you might be disagreeing on straightforward issues of non-moral fact: what the consequences of a certain policy might be, what others think, how they will be affected, and so on. But suppose you have come to agreement on all of that. What is left? Stevenson says that there is just 'disagreement in interest'. You have different attitudes, and that is the end of it. There is no mistake that the other person is making.

A second issue concerns moral reasoning. Suppose you are worrying about some moral issue: deciding what is the right thing to do. Again, assume that you know all the relevant non-moral facts; but you are still unsure about what you should do. What can the expressivist say that you are deliberating about? They might say that you are deliberating about how to feel. Does that capture the idea? Of course you want to feel the right way. But you might think that you should feel a certain way towards an action because it is right. Whereas the expressivist will have to say that you judge an action is right because you feel a certain way.

Stevenson's response to these worries is to deny that they really have any substance:

And now, have I really pointed out the "vital" sense of "good"?

I suppose that many will still say "No", claiming ... that my analysis, like all others given in terms of interest, is a way of begging the issue. They will say: "When we ask 'Is X good?' we don't want mere influence, mere advice. We decidedly don't want to be influenced through persuasion, nor are we fully content when the influence is supported by a wide scientific knowledge of X. The answer to our question will, of course, modify our interests. But this is only because a unique sort of *truth* will be revealed to us—a truth which must be apprehended *a priori*. We want our interests to be guided by this truth, and by nothing else. To substitute for such a truth mere emotive meaning and suggestion is to conceal from us the very object of our search."

I can only answer that I do not understand. What is this truth to be *about*? For I recollect no Platonic Idea, nor do I know what to *try* to recollect. I find no indefinable property, nor do I know what to look for. And the "self-evident" deliverances of reason, which so many philosophers have claimed, seem, on examination, to be deliverances of their respective reasons only (if of anyone's) and not of mine.

It's somewhat contentious to say that the truth must be apprehended *a priori*; moral realists might deny that. But still, Stevenson effectively challenges the realist to substantiate their claim that there must be something more. It is not easy to break the deadlock here. Instead we'll focus on internal problems for expressivism.

2. Can the expressivist say that moral sentences are neither true nor false?

The issue here is not a problem for expressivism per se; but rather an issue about the compatibility of expressivism with another popular view. In *Language, Truth and Logic*, Ayer held that a sentence like *Stealing money is wrong* 'expresses no proposition which can be either true or false'. But earlier in the same book he had embraced a disquotational or minimalist or redundancy view of truth, according to which, to assert that a sentence is true is equivalent to simply asserting that sentence itself (see pp. 107 and 87ff. respectively). But it is not obvious that those two doctrines can be held together.

Let's first say a little more about the disquotational view of truth. According to this view, first clearly stated by Ramsey, the truth predicate basically serves to take the quotation marks off (to 'disquote'): I can say '*Snow is white*' is true; or, equivalently I can just say *Snow is white*. (Note that I'm also using italics as a device of quotation here, so that we don't get swamped with quotation marks!) Likewise the falsity predicate is used in place of the negation of the sentence. If that is all there is to the truth predicate you might wonder why we have it; the disquotationalist's answer is that it is used in cases where we are in no position to assert the sentence on its own. For instance, I might say, *Whatever she told him is true*, even though I don't know what she said. On a standard disquotationalist understanding, that enables us to say something that would otherwise require an infinite disjunction: Either she said that p, and p; or she said that q, and q ... and so on for all the things she might possibly have said.

This is a nice simple account of truth. It doesn't require worrying about what many of the issues that other accounts have had; but it doesn't seem to be compatible with expressivism as Ayer defended it. For if the expressivist wants to say that the sentence *Eating meat is wrong* has a perfectly legitimate use, albeit an expressive one, then the sentence '*Eating meat is wrong*' is true will also have a perfectly legitimate use, namely exactly the same one. But then how can we insist that *Eating meat is wrong* is neither true nor false?

Ayer seems not to have noticed this problem. Stevenson, in his 1944 book saw it very clearly. His solution was to link expressivism inescapably to noncognitivism. The reason that we don't want to say that moral sentences express propositions isn't that they are neither true nor false; it is that they do not express beliefs. So although expressivism is still a thesis about the status of sentences in a public language, the distinctive claim about them derives from the psychological states that they express.

3. Indirect uses and the Frege-Geach problem

Recall that Stevenson focussed on the use of ethical sentences to recommend; Ayer focussed on their use to express an attitude. But very soon critics started to point out that ethical sentences can be used in many other ways. So Ross, in his 1935/6 Gifford lectures, published in 1939, observed that they can be used in a wide range of cases, for instance in conditionals:

There is no doubt that such words as 'you ought to do so-and-so' may be used as one's means of so inducing a person to behave a certain way. But if we are to do justice to the meaning of 'right' or 'ought', we must take account also of such modes of speech as 'he ought to do so-and-so', 'you ought to have done so-and-so', 'if this and that were the case, you ought to have done so-and-so', 'if this and that were the case, you ought to do so-and-so', 'I ought to do so-and-so.' Where the judgement of obligation has referenced either a third person, not the person addressed, or to the past, or to an unfulfilled past condition, or to a future treated as merely possible, or to the speaker himself, there is no plausibility in describing the judgement as

command. But it is easy to see that ‘ought’ means the same in all these cases, and that if in some of them it does not express a command, it does not do so in any.
Ross, *Foundations of Ethics*, pp. 33–4

The problem arises even for negation. If *Eating meat is wrong* expresses my disapproval of eating meat, what does *Eating meat is not wrong* express? Presumably not my approval of eating meat — I could think it wasn’t wrong without thinking it was right. Nor does it express my failure to either approve of eating meat or to disapprove of it, for it is quite compatible with *Eating meat is right*.

One approach the expressivist might take is to claim that moral terms in asserted contexts mean something different to what they mean in the scope of negations, conditionals etc. Ross simply denies this, but he doesn’t give an argument for it. But Peter Geach, developing what he called ‘the Frege point’ — that words mean the same whether they are asserted or not — did. Geach asks us to consider the following:

If doing a thing is bad, getting your little brother to do it is bad.
Tormenting the cat is bad.
Ergo, getting your little brother to torment the cat is bad.
Peter Geach ‘Assertion’ 1965 p. 463

That certainly looks like a *valid* argument, whether or not you believe the premises. But if it is valid the term ‘bad’ had better mean the same in each occurrence. This objection to expressivism has become known as the Frege-Geach problem.

4. Three Responses to the Frege-Geach Problem

i Attitudes to attitudes

A first response — or rather set of responses, since he refined the details several times — came from Simon Blackburn. His basic thought was that the conditional P2 could be seen as an evaluative attitude towards other attitudes: one disapproves of: disapproving of tormenting the cat in general, while not disapproving of it when done by one’s little brother. And then the argument gets a kind of informal validity:

Anyone holding this pair [expressed by P1 and P2] must hold the consequential disapproval: he is committed to disapproving of getting little brother to [torment the cat], for if he does not his attitudes clash. He has a fractured sensibility which cannot itself be an object of approval. (Blackburn *Spreading the Word*, 195)

This is perhaps plausible in this case: P2 does look like a moral claim itself. But how general is the approach? Can it be extended to other embedded sentences?

It’s unclear whether it even extends to sentences under negation. We saw last week that that poses problems for the expressivist. Consider again:

(I) Eating meat is not wrong

If we want to understand this as involving higher order attitudes, then a plausible candidate is that of disapproving of disapproving of eating meat. Does that give us what we want? After all, (I) should presumably be inconsistent with

(2) Eating meat is wrong

But can't I disapprove of something as in (2), and also disapprove of my own disapproval as in (1)? That might show a certain degree of self-loathing, but it doesn't seem incoherent: I might, for instance, think that I am in general much too ready to disapprove of things; it is a fault in myself, of which I heartily disapprove. (Compare Frankfurt on higher-order desires.) So (1) and (2) aren't obviously inconsistent.

This raises the broader problem of whether the requirement of avoiding a 'fractured sensibility' is really enough to underpin the validity of arguments containing moral terms. There is much discussion of what is wrong with what has come to be known as Moore's paradox:

(3) It is raining but I don't believe that it is.

But almost everyone agrees that while that sentence is somehow self-defeating, it is not contradictory. After all, the third person version will frequently be true:

(4) It is raining but Imelda doesn't believe that it is.

Valid arguments are typically understood as those whose premises are inconsistent with the negations of their premises. But the kind of inconsistency that Blackburn's approach imputes here is more like the inconsistency involved in Moore's paradox, and that isn't really inconsistency at all. So the corresponding notion of validity isn't really validity.

ii Hybrid Theories

To get a grip on this approach, start by considering slurs. Philosophers are understandably a bit sensitive about using active ones, so often the focus is on something obsolete, like 'boche', a French/English WWI derogatory term for a German. If someone had said, in 1916:

(5) Frege is a confounded boche; I'm not reading his work any more

then presumably they would have said something descriptive by the first part of that sentence (Frege is German), but they would also have expressed their attitude (contempt or whatever) towards him. Perhaps moral terms could be like that, combining both (i) a descriptive element, that explains how they function in deductive arguments etc., and so avoiding the Frege-Geach objection; and (ii) an expressive element that explains the motivational internalism. It might seem that this is bolstered by a hybrid approach to the mental states that moral sentences express, one that combines cognitive and affective elements.

However, getting the details right has proved difficult, and there are many competing approaches.

A first question is: What is the descriptive content meant to be? In the case of the first half of (5) it is straightforward: it is that Frege is a German. But in the case of

(2) Eating meat is wrong

it is not so obvious what it is. Some have suggested that a better parallel is with derogatory terms like 'jerk', which are less obviously descriptive. But even that does seem to have some descriptive content about how the person acts, and, in consequence, it is not clear that it is bound to be derogatory:

(6) He's a jerk, but actually I quite like jerks.

Perhaps this provides a good model for 'thick' ethical terms like 'brave', 'loyal' etc.; but we are still looking for a model for the thin ones like 'good', 'bad', 'right', 'wrong', 'permissible'.

A second issue concerns whether we should think of the expressive aspect of moral terms as providing a component of the meaning in the way that the hybrid theory claims. If someone speaks with a certain class or regional accent they give away information about themselves (hard to avoid in the UK). But presumably we don't want to think that that information is part of the meaning of what they say. Likewise, if I write about a car bonnet in the US, rather than a hood, I give away the information that I'm not American. But again that information is not part of the meaning of what I say. Presumably what is special about moral terms is that the expressive meaning is conventionally associated with it (recall Stevenson). But might it not be a conventional implicature rather than part of the content?

A third issue concerns the link between the descriptive content and the emotive. Presumably they are not simply conjoined: it is not that it is good, *and* I'm moved. The descriptive content meant to be the reason that people are moved.

None of these are obviously unanswerable concerns, but there is currently not much agreement from proponents of this approach on how to respond to them.

iii Systematic new contents

The main figure here is Gibbard. Unlike Blackburn's higher order account, Gibbard's account is completely general: it can accommodate *any* expressive sentence. It works by using exactly the same machinery that is standardly used to handle belief ascriptions to handle non-cognitive attitudes. And because of this, it can handle not just sentences that combine two non-cognitive attitudes (the kind that featured in our original formulation of the Frege-Geach problem) but also contents that combine non-cognitive attitudes and attitudes of belief (e.g. 'If he's in trouble you should help him'), or indeed any other attitude provided it has the right kind of content (we'll see what this restriction means at the end).

To see how it works, we start with one of the standard ways of representing belief. The contents of beliefs can be thought of as sets of possible worlds: the belief that *p* is the set of worlds in which *p* is true. Equivalently, it is the set of worlds that you arrive at when all the worlds in which *p* is not true are excluded. Adding further beliefs can then be thought of as the exclusion of further worlds. So if I believe that *p* and I believe that *q*, my belief state is represented by the class of worlds that remains when I exclude all of the worlds in which *p* is false, and all of the worlds in which *q* is false. An agent's total belief state can then be thought of as the set of worlds that is left when this process is done with all of their beliefs; equivalently, it is the set of worlds that are compatible with all of their beliefs. Call these the agent's *belief worlds*.

This approach gives a clear account of the sentential connectives, and also explains why the classical argument forms are valid. So, for instance, an agent believes If *p* then *q*, just in case every *p*-world amongst their belief worlds is also a *q*-world; and modus ponens is valid for belief given this. Parallel moves work for the other sentential connectives. (The account isn't without its critics: if *p* and *q* are true in the same set of worlds—for instance, any two necessary truths—then they seem to have the same content on this approach. A great deal of work has gone into softening the threat of this, but we won't pursue that here.)

Fundamental to Gibbard's approach is the realization that exactly the same machinery can be used to describe attitudes other than belief. Suppose that you accept a set of norms, such as the norm against lying. And suppose, in a standard expressivist way, you think that moral sentences express those norms, so that the sentence 'lying is wrong' expresses the norm against lying, but makes no assertion. Then, even though the attitudes expressed here are non-cognitivist ones, for each norm we can *divide the worlds into those that the agent takes to conform to that norm, and those that do not*. So, for instance, we can divide the worlds into those in which no one lies, thereby conforming to the lying norm, and those in which someone does. And then, just as with beliefs, we can further restrict the class of worlds to those that conform to all of the agent's norms.

Because we are treating norms in the same way that we treated beliefs — as a way of partitioning worlds — we can understand operations on them in just the same way. So take the conditional 'If lying is wrong, then encouraging others to lie is also wrong'. That sentence is true for the agent just in case the worlds that they take to conform to the norm expressed by the sentence 'lying is wrong', are also worlds that they take to conform to the norm expressed by the sentence 'encouraging others to lie is wrong'. (It is, of course, debatable whether that is true—some might understand those norms as having nothing to do with each other.) So, even though the atomic, norm-expressing sentences get a non-cognitive content, the conditional in which they occur does not. And we can see that modus ponens is valid for it. If the speaker endorses the norm expressed in the antecedent, and on their understanding of the norms, the conditional is true, then they will endorse the norm expressed by the consequent.

Moreover, we can make sense of conditionals combining sentences that express beliefs and sentences that express norms. Take the norm expressed by the sentence 'You should help people who need it'. A natural (though not uncontroversial) reading of this is as a conditional, with a descriptive antecedent and a norm expressing consequent: 'If someone needs help, you should help them'. On the account we are considering, this will be true given your norms when applied to some individual, A, just in case all of your belief worlds in which A needs help are worlds in which you would embrace the norm 'I should help A', i.e. worlds that would only conform to that norm if you did help A. And so again such a mixed conditional validates modus ponens.

The approach described here is very general. In the characterization I have given, following Gibbard's early work, we have seen how it can combine belief expressing sentences with *norm* expressing sentences. In later work Gibbard focuses on sentences that express *plans*. There is no obvious end to the different attitudes that could be accommodated. All they need is to have a structure that enables us to partition the worlds in accordance with whether they do or do not conform to them; to that extent they have a broadly representational content. Not all sentences have this content. Dreier supposes that instead of greeting Bob with a simple 'Hi', we might say, in greeting, 'Bob is hiyo'. Grammatically that move is easy enough to make, but in making we haven't arrived at a sentence with which the world either does or doesn't conform; so it will make no sense to put it into a conditional.