NOÛS 42:4 (2008) 573–599

How Expressivists Can and Should Solve Their Problem with Negation

MARK SCHROEDER
University of Southern California

Expressivists have a problem with negation. The problem is that they have not, to date, been able to explain why 'murdering is wrong' and 'murdering is not wrong' are inconsistent sentences. In this paper, I explain the nature of the problem, and why the best efforts of Gibbard, Dreier, and Horgan and Timmons don't solve it. Then I show how to diagnose where the problem comes from, and consequently how it *is* possible for expressivists to solve it. Expressivists should accept this solution, I argue, because it is demonstrably the *only way* of avoiding the problem, and because it *generalizes*.

Once we see how to solve the negation problem, I show, it becomes easy to state a constructive, compositional expressivist semantics for a purely normative language with the expressive power of propositional logic, in which we can for the first time give explanatory, formally adequate expressivist accounts of logical inconsistency, logical entailment, and logical validity. As a corollary, I give what I take to be the first real expressivist explanation of why Geach's original moral *modus ponens* argument is genuinely logically valid. This proves that the problem with expressivism *cannot be* that it can't account for the logical properties of complex normative sentences. But it does not show that the same solution can work for a language with both normative and descriptive predicates, let alone that expressivists are able to deal with more complex linguistic constructions like tense, modals, or even quantifiers. In the final section, I show what kind of constraints the solution offered here would place expressivists under, in answering these further questions.

© 2008, Copyright the Authors
Journal compilation © 2008, Wiley Periodicals, Inc.

1.1 Introduction

Famously, metaethical expressivists have a problem accounting for the meanings of normative terms in embedded contexts. Geach (1960), (1965) and Searle (1962) originally contended that non-cognitivist views were inconsistent with the obvious datum that 'murdering is wrong' means the same thing when embedded in 'it is not the case that murdering is wrong' as when standing alone, and offered as evidence of this datum the fact that the two sentences are logically inconsistent, a fact that they said could only be due to the fact that 'murdering is wrong' has the same meaning in both places.

Geach and Searle understood non-cognitivists to be giving a speech-act criterion for what it was for a string of words to mean what 'murdering is wrong' does. But they noticed that someone who asserts 'murdering is not wrong' is not engaging in the same speech act as someone who asserts 'murdering is wrong'. So that is why they concluded that by non-cognitivist lights, the words in those sentences must have different meanings. This was the original embedding problem for expressivism—the Frege-Geach Problem.

Following Hare (1970), however, expressivists have held that this misinterprets their view. It is true that 'murdering is wrong' and 'murdering is not wrong' express different attitudes, they say. But that does not show that the words in the second have a different meaning from the words in the first. Similarly, expressivists allege, 'grass is green' and 'grass is not green' have different truth-conditions. But that doesn't show, on a truth-conditional semantics, that the words in the second have a different meaning than the words in the first. So expressivists like Blackburn, Gibbard, and Horgan and Timmons have followed Hare in contending that what expressivists need is a *compositional semantics*. The view that the meaning of 'murdering is wrong' is that it expresses disapproval of murdering, they say, is not a criterion for what it takes for a string of letters to have the same meaning as 'murdering is wrong'. Rather, it is simply a base clause in a recursive compositional semantics which will ultimately aspire to tell us the meaning of every complex normative sentence, by telling us what mental state it expresses as a function of the mental states expressed by its parts.

So, Hare (1970), Blackburn (1973), (1984), (1988), (1998) and Gibbard (1990), (2003) have argued, Geach and Searle were wrong—expressivism is *not* inconsistent with the claim that normative sentences have the same meaning when embedded. They can explain this by appealing to the claim that the meaning of the complex sentence is a function of the meanings of its parts.[1] But this does not mean that compositional semantics comes to expressivists for free. On the contrary, all of the same data that Geach and Searle pointed to as evidence that normative terms mean the same thing when embedded as unembedded all come back, as data that need to be explained by

an adequate expressivist account of what attitudes are expressed by complex sentences. The idea here is simple. Not just any account of the meaning of 'murdering is not wrong' as a function of the meaning of 'murdering is wrong' will be adequate. Only one that allows us to explain why the two sentences are inconsistent will be adequate. It is due to the meaning of 'not' that 'murdering is wrong' and 'murdering is not wrong' are inconsistent. So an account of the meaning of 'not' owes us an explanation of why this is so.

This is the embedding problem on which expressivists and their opponents have focused their attention since 1970.[2] The challenge is this: give a compositional account of the attitudes expressed by complex normative sentences as a function of the attitudes expressed by their parts. And then expressivists must explain why this account predicts that sentences with that structure have the right kinds of properties: for example, that negations are inconsistent with the sentences they negate, that conditionals can be used in *modus ponens*, that disjunctions validate disjunctive syllogism, and so on.

1.2 Inconsistency

Some properties of complex sentences will be harder to explain than others. For example, explaining the properties of questions might require a better understanding of how ordinary descriptive questions work than is suitably uncontroversial. It's also easy to see that explaining the validity of simple argument forms like *modus ponens* and disjunctive syllogism ought to be harder than explaining the inconsistency of atomic sentences and their negations. After all, it is necessary to explain the logical validity of the argument P,P⊃Q;Q to show that {P,P⊃Q,∼Q} is a logically inconsistent set.[3] But first, understanding this requires having a semantics for '⊃', which simply adds extra complication to the investigation. Moreover, surely understanding *logical* inconsistency requires understanding inconsistency in the first place, understanding pairwise inconsistency ought to be easier than understanding inconsistency of larger sets, and understanding the inconsistency of atomic sentences and their negations ought to be the easiest of all. So by all counts, one of the easiest things for expressivists to explain, and methodologically one of the things they should focus on *first*, has got to be why 'murdering is wrong' and 'murdering is not wrong' are inconsistent.

It is a major embarrassment for expressivism, then, that no one has ever shown how to do this! And it is certainly not that no one has tried. We'll see in part 2 why it looks like such a hard thing for expressivists to explain, and in part 3 why all existing accounts fail to explain it. But first, we have to look at what tools expressivists have, with which to explain it.

According to expressivism, the meaning of 'murdering is wrong' is given by the mental state that it expresses. Expressivists hold that this is a non-cognitive attitude rather than a belief, and I stipulatively call it 'disapproval of murdering'. Similarly, expressivists hold that 'murdering is not wrong' will

have a meaning that is given by the attitude it expresses. So any explanation of the inconsistency of 'murdering is wrong' and 'murdering is not wrong' will have to proceed by way of properties of these two mental states.

The first thing that expressivists noted about this problem was that they could solve it in the case of ordinary descriptive sentences. An ordinary explanation of why 'grass is green' and 'grass is not green' are inconsistent would start by noticing that they have truth-conditions which cannot simultaneously be satisfied. Obviously expressivists have a difficulty with generalizing this account to normative sentences, however. So they asked, 'what other account could we give of the inconsistency of descriptive sentences that *could* be generalizable?' And the answer was that they could explain the inconsistency of 'grass is green' and 'grass is not green' by appeal to the inconsistency of the *beliefs* that are expressed by these two sentences—by appeal to inconsistency in mental states, rather than inconsistency in sentential truth-conditions.

On this view, 'grass is green' expresses the belief that grass is green, and 'grass is not green' expresses the belief that grass is not green. But it is inconsistent to have both of these beliefs. In Gibbard's terminology, if you have one of these beliefs, then you *disagree* with someone who has the other. So expressivists say that *this* is what makes the two sentences inconsistent, and claim that it will be enough to explain the inconsistency of 'murdering is wrong' and 'murdering is not wrong' to explain why the mental states expressed by them *disagree* in the same kind of way as these beliefs do. All of this is consistent with going on to say that beliefs are inconsistent because they have truth-conditions that are inconsistent; in fact, it is precisely this which guarantees that in the case of descriptive sentences, the expressivist account of inconsistency in terms of inconsistency of attitudes is no less formally adequate than a direct account of inconsistency in terms of inconsistency of truth-conditions. So long as a descriptive sentence expresses a belief with its same truth-conditions, and beliefs are inconsistent when their truth-conditions are, the two accounts get the same results. Yet the account in terms of attitudes at least has the right shape to generalize to the case of normative sentences.

1.3 Inconsistency-Transmitting Attitudes

Of course, there are a number of obstacles to this generalization. The first is the worry that it might turn out that beliefs are really the only kinds of mental state that 'disagree' with one another, in Gibbard's sense—which can be inconsistent with one another in the way that beliefs are. Expressivists, however, have always held that intentions offer a good model for a non-cognitive state that is subject to the same kind of inconsistency. Just as it is a mistake to believe p and believe $\sim p$, expressivists have pointed out that it is a mistake to intend that p and also intend that $\sim p$, and you are in

disagreement with someone, if you intend that p and she intends that $\sim p$. Yet intention seems like a non-cognitive attitude. At the least, it comes from the practical, world-to-mind side of our psychologies, rather than from the theoretical, mind-to-world side. So Stevenson (1937) used intention in order to argue that practical disagreements are possible, and Gibbard (2003) uses it as his model, in trying to argue that expressivism *must* be a workable view.

So if intentions are an example of a non-cognitive attitude subject to the same kinds of inconsistency as beliefs are, then maybe there are others, including perhaps disapproval. Or maybe the non-cognitive states expressed by normative sentences *are* intentions or states involving intentions, as Gibbard (2003) suggests. Yet expressivists face yet another obstacle. For many philosophers have argued that the inconsistency between intentions is itself explained by inconsistency between beliefs and the fact that intention involves belief.[4] Michael Bratman (1993) calls this view *cognitivism about instrumental reason* and has offered a number of interesting and I think powerful arguments against it. Still, if cognitivism about instrumental reason turned out to be true, expressivists would be without a good model of a noncognitive attitude that is subject to the right kinds of inconsistency for an expressivist explanation of inconsistency between sentences to work.

Still, assuming that there is no such problem, we need to draw one more important observation about *in what cases* beliefs and intentions are inconsistent with (disagree with) one another. Beliefs are inconsistent with one another when their *contents* are inconsistent. Likewise, intentions are inconsistent with one another when their *contents* are inconsistent. The sense in which they are inconsistent is not merely *that* their contents are inconsistent. It is not inconsistent to both wonder whether p and wonder whether $\sim p$, but those are states with inconsistent contents. Still, all of the good paradigms that we have of what Gibbard calls disagreement in attitude arise in cases of the same attitude toward inconsistent contents. Let us call attitudes like belief and intention, which have this property, *inconsistency-transmitting attitudes*:

> **inconsistency-transmitting:** An attitude A is inconsistency transmitting just in case bearing A toward inconsistent contents is inconsistent.

Provided that cognitivism about instrumental reason is wrong, then, expressivists have a good model, in intention, for a non-cognitive attitude that is inconsistency-transmitting. Inconsistency-transmittingness, therefore, is the kind of feature to which expressivists are intelligibly entitled to appeal in their explanations. It is intelligible for expressivists to hope that whatever explains the inconsistency-transmitting character of belief and intention will also explain why disapproval is inconsistency-transmitting.

If so, then expressivists could explain why some normative sentences are inconsistent with one another. They could explain why 'murdering is wrong', which expresses disapproval of murdering, is inconsistent with the sentence

that expresses disapproval of not murdering, by appeal to the assumptions that murdering and not murdering are inconsistent, and that disapproval is an inconsistency-transmitting attitude, together with the account that two sentences are inconsistent just in case the attitudes that they express are. So this is an expressivist-respectable account of at least one case of inconsistency.

2.1 The Negation Problem

Unfortunately, however, as Nicholas Unwin (1999), (2001) has pointed out forcefully, this does not yield expressivists an explanation of why atomic normative sentences are inconsistent with their negations. This is because the sentence that expresses disapproval of not murdering is 'not murdering is wrong'. But the negation of 'murdering is wrong' is 'murdering is not wrong'.

In fact, Unwin argues, the problem is a deep one. There are three places to insert a negation in 'Jon thinks that murdering is wrong', all of which receive distinct semantic interpretations:

w Jon thinks that murdering is wrong.
n1 Jon does not think that murdering is wrong.
n2 Jon thinks that murdering is not wrong.
n3 Jon thinks that not murdering is wrong.

Sentence n1 denies Jon the view that murder is wrong, n2 attributes to Jon a negative view about the wrongness of murdering, and n3 attributes to Jon a positive view about the wrongness of not murdering. According to n2 he thinks that murdering is permissible, whereas according to n3 he thinks that it is obligatory. Conflating any two of these three would be a disaster.

Yet that is precisely the danger for expressivists. For according to expressivism, thinking that murdering is wrong is being in the mental state expressed by 'murdering is wrong'. That is, it is disapproving of murdering. But there are simply not enough places to insert a negation in 'Jon disapproves of murdering' to go around:

w* Jon disapproves of murdering.
n1* Jon does not disapprove of murdering.
n2* ???
n3* Jon disapproves of not murdering.

There is simply one place not enough for the negations to go around. There is no way to account for the meaning of n2 by applying 'not' somewhere to the meaning of w. And that makes it look very much like expressivists are

not going to be able to offer a satisfactory explanation of why 'murdering is wrong' and 'murdering is not wrong' are inconsistent.

The problem is this. If the inconsistency of 'murdering is wrong' and 'murdering is not wrong' is to be explained, then these two sentences must express inconsistent states of mind. And if the inconsistency of these states of mind is to be explained by means of assumptions that it is respectable for expressivists to make, then it must be explained in terms of the inconsistency-transmittingness of some attitude—they must be the same attitude toward inconsistent contents. So since 'murdering is wrong' expresses a state of disapproval of murdering, it follows that 'murdering is not wrong' must also express a state of disapproval. It must be disapproval of x, for some value of x that is inconsistent with murdering.

But it is easy to prove that there is no such possible value of x. To see why, compare the following four sentences:

1	Stealing is wrong.	→	DISAPPROVAL(stealing)
2	Stealing is not wrong.	→	DISAPPROVAL(x)
3	Not stealing is wrong.	→	DISAPPROVAL(not stealing)
4	Not stealing is not wrong.	→	DISAPPROVAL(y)

1 and 2 are inconsistent sentences, as are 3 and 4. So if their inconsistency is to be explained in terms of disagreement between the mental states that they express—states which rationally conflict with each other in just the same way that beliefs with inconsistent contents do—and this is to be explained by the fact that disapproval, like belief and intention, is the sort of attitude that it rationally conflicts in this way to hold toward inconsistent contents, then 2 and 4 must express some states of disapproval. 2 must express disapproval of something inconsistent with stealing, in order to explain why 1 and 2 are inconsistent, and 4 must express disapproval of something inconsistent with not stealing, in order to explain why 3 and 4 are inconsistent. But if x is inconsistent with stealing, and y is inconsistent with not stealing, then it follows that x and y must be inconsistent with each other. But this yields the prediction that the states of mind expressed by 2 and 4 rationally conflict in exactly the way required in order to explain the inconsistency of 2 and 4. But 2c and 4c are not inconsistent sentences!

So it follows that 'murdering is not wrong' cannot express disapproval of anything at all, if things are going to turn out alright. It must express another state—call it *tolerance of murdering*. But if so, then its inconsistency with 'murdering is wrong' cannot be explained by the assumption that disapproval is inconsistency-transmitting! That is, it cannot be explained by appeal to the kinds of assumptions that it is respectable for expressivists to make.

Where does this leave current expressivist views? Appealing to a distinct attitude of tolerance in order to provide a semantics for the negations of

'wrong' sentences is a natural idea, and it is essentially shared by all existing expressivist views. After all, it is an old observation that 'permissible', 'impermissible', 'obligatory', and 'unobligatory' can all be interdefined using negation. But the negations have to appear in *two* places, both 'inside' and 'outside' the term we are using to define the others. For example, 'permissible' is 'not impermissible'. So given that we understand 'external' negation, we only need one of those four in order to define the others. But the problem is that 'external' negation is precisely what expressivists are trying to give an account of. The whole point of solving the embedding problem for the case of negation, recall, is to say what mental state is expressed by the negation of normative sentences. So rather than define 'permissible' as 'not impermissible', they define 'not impermissible' as meaning, 'permissible', and 'not permissible' as meaning, 'impermissible'. And then they say that 'impermissible' sentences express an attitude called *disapproval* and that 'permissible' sentences express an attitude called *tolerance*, and assume that it is inconsistent to have both of these attitudes toward the same thing—that someone who does so *disagrees* with herself, in Gibbard's sense.

This is a perfectly good formal move. Out of 'permissible', 'impermissible', and sentential negation, any two of the three will allow us to introduce the third in a way that is formally adequate. The usual practice is to take sentential negation as understood and interdefine the other two, but it is *open*, in some sense, for expressivists to do things the other way around. To make this kind of move, they need at least one attitude from the {'permissible','unobligatory'} pair *and* at least one from the {'impermissible','obligatory'} pair, *both* of which they take as primitive, in order to define external negation. Blackburn's (1988) account used the attitudes he called 'tolerance' and 'hooraying', corresponding to 'permissible' and 'obligatory'. (His earlier (1984) discussion wrongly focused on 'booing' and 'hooraying', which corresponded to 'impermissible' and 'obligatory', falling directly into the trap mentioned in the first paragraph of this section.) I'm focusing on disapproval and tolerance, corresponding to 'impermissible' and 'permissible'. It doesn't matter, really, which we choose, but the point is that we need one from each pair in order to be able to define 'external' negation.

2.2 The Need For an Explanation

Let me be clear that we *can* do things this way. So the problem is not that expressivists have no answer as to what n2 means. The answer is that it means that Jon tolerates murdering. But once we do things in this way, it should be very clear that we have left completely unexplained and apparently inexplicable why 'murdering is wrong' and 'murdering is not wrong' are inconsistent.

Suppose, for example, that someone tells you that when she uses the word 'not' or the prefix 'im-' immediately before 'permissible', they are not to be understood as meaning what 'not' normally does. Instead, she says, she believes in distinct, unanalyzable, and non-interdefinable properties of permissibility and impermissibility. And then suppose that she tells you that she also believes that it is impossible—*logically* impossible—for something to be both permissible and impermissible. Finally, she tells you, by 'not permissible' she means 'impermissible' and by 'not impermissible', she means 'permissible'. That is why, she tells you, 'murdering is permissible' and 'murdering is not permissible' are logically inconsistent sentences. It is because the latter means 'murdering is impermissible', and permissibility and impermissibility are assumed to be logically incompatible.

Surely this account leaves something to be explained! Obviously her view will be a bad view about permissibility and impermissibiliy unless they *do* turn out to be incompatible. But that does not mean that she is entitled to assume it! On the contrary, her view seems to have written out of existence everything that could be used to explain why permissibility and impermissibility are incompatible, and given us an account of why this sentence and its negation are inconsistent that appears to have nothing to do with the meaning of 'not', into the bargain. Expressivists are in the same position with respect to disapproval and tolerance. The negation problem shows that they can't simply be interdefined, which leads to the conclusion that they are distinct and unanalyzable attitudes. But if they are, then why on earth is it inconsistent to hold them toward the same thing?

One more observation is requisite in order to draw out exactly how difficult this problem is for expressivists, and to understand how inadequate existing answers are. The observation is to compare just how different this kind of inconsistency would be, between disapproval of murdering and tolerance of murder, from the familiar kinds of inconsistency for which expressivists have good models elsewhere. All of the other good models of inconsistency between mental states arose in the case of inconsistency-transmitting attitudes. They were all cases of the same attitude toward inconsistent contents. Call this *A*-type inconsistency. *A*-type inconsistency is relatively easy to explain, because to explain it all that you need is a general fact about an attitude type (that it is inconsistency-transmitting) and an easy claim about their contents (that they are inconsistent). But tolerance of murdering and disapproval of murdering are two *distinct* and apparently *logically unrelated* attitudes toward the *same* content. Call this *B*-type inconsistency. *A*-type inconsistency is something that we should all recognize and be familiar with. It happens with beliefs, for example. But *B*-type inconsistency is not something that expressivists should be taking for granted, because there are few good examples of it. Assuming that disapproval and tolerance of murdering are inconsistent is taking for granted everything that expressivists need to explain.

The problem, moreover—and it is hard to overemphasize this point—only gets harder when we start considering how to negate complex normative sentences. If an atomic sentence like 'murdering is wrong' expresses disapproval of something, it is not clear exactly what mental state should be expressed by 'murdering is wrong and stealing is wrong', but by the same reasoning as for negation, we can't define it out of disapproval, for we again get three distinct ways of inserting a conjunction:

&1 Jon thinks that murdering is wrong and Jon thinks that stealing is wrong.

&2 Jon thinks that murdering is wrong and stealing is wrong.

&3 Jon thinks that murdering and stealing is wrong.

Since belief does not agglomerate across conjunction (you don't believe the conjunction of everything that you believe), we shouldn't collapse &1 and &2. And clearly both are distinct from &3. So the attitude expressed by 'murdering is wrong and stealing is wrong' will turn out to be distinct from disapproval as well. And once it is, it's obvious that the attitude expressed by its negation can't turn out to be just an ordinary state of tolerance, either.

If that doesn't seem like too many attitude-kinds yet, then try taking the conjunction of 'murdering is wrong' with this negation. And then try negating that. It's a good exercise to see just how quickly we end up needing to posit an infinite list of distinct kinds of attitude to go along with disapproval and tolerance—for every pair of which there will be inconsistency relations for which we have no explanation. Expressivists thus get themselves into not just *one* appeal to brute *B*-type inconsistency relationship between attitudes, but to *infinitely many* brute *B*-type inconsistency relationships between attitudes. So even if there really are some *B*-type inconsistency relationships that cannot be further explained (such as those between believing that *p* and doubting that *p*, or between intending to do *A* and believing that you won't do A[5]), expressivists who adopt this kind of picture will be in a much worse way—needing to postulate infinitely many such brute *B*-type inconsistency relationships. Surely all of these leave something to be explained.

And that's just for one normative predicate, 'wrong'. Whereas for all descriptive predicates put together (and there are a lot of them), there is only one basic attitude-kind: belief. If the view on which every complex construction yields a distinct attitude-kind sounds simply too incredible to you to be worth going on, it's worth noting not only that this is essentially a commitment of all existing expressivist views, but that it has recently been explicitly defended in print.

3.1 Horgan and Timmons

The culprits are Terry Horgan and Mark Timmons (2006), who offer an 'expressivist logic', which they claim shows how to solve the embedding

problem. They postulate that there is a basic non-cognitive attitude, which they call 'ought-commitment'. There is also a basic cognitive attitude, which they call 'is-commitment' and other expressivists would simply call 'belief' (Horgan and Timmons make a big deal out of the fact that they get to call non-cognitive attitudes 'beliefs', too, so they resist this characterization).

The details aren't crucial, but in Horgan and Timmons's official regimented language, they have 'non-sentential formulas', which are like sentences from ordinary predicate logic and represent the contents of mental states, and then they have 'sentential-formula-forming operators', which correspond to types of mental state. For example the basic sentential-formula-forming operators are **I[ ]** and **O[ ]**, which represent is-commitment and ought-commitment. The 'sentences' of Horgan and Timmons' regimented expressivist language result from applying sentential-formula-forming operators to non-sentential formulas. We can think of them like descriptive names for the mental states that the sentences express; each sentence has exactly one sentential-formula-forming operator taking widest scope, which corresponds to the kind of attitude expressed by the sentence. The subsentential formulas filled in to its gaps tell us the content of that state. For example, if 'Ws' is a subsentential formula meaning that snow is white (think 'W' = 'white', 's' = 'snow'), then '**I[**Ws**]**' is a sentence. It expresses is-commitment to snow being white (the belief that snow is white). '**O[**Ws**]**' is also a sentence. It expresses ought-commitment to snow being white.

The way that Horgan and Timmons deal with complex sentences of all kinds is simple. Starting with negation they tell us that for any sentential-formula-forming operator **Ω**, there is a distinct sentential-formula-forming operator, **¬Ω**. Since each sentential-formula-forming operator corresponds to a kind of mental state, starting with **I[ ]** corresponding to is-commitment and **O[ ]** corresponding to ought-commitment, all of this is just a complicated way of saying that whenever you negate a sentence, it expresses a different kind of attitude than the sentence that you negated. They give corresponding stories about conjunction, disjunction, the material conditional and biconditional, and the unary existential and universal quantifiers. Every time you take one or more sentences and make a more complex sentence, according to Horgan and Timmons, the complex sentence expresses a new and distinct kind of mental state.

Horgan and Timmons call these 'logically complex commitment states', and define them in terms of their inferential role. That is, what they tell us about the state corresponding to **¬O[ ]**, is that bearing it to some content is inconsistent with bearing **O[ ]** toward that content. So in essence Horgan and Timmons' view amounts to the hypothesis that there is an unfathomably huge hierarchy of distinct kinds of mental state, together with unsupported confidence that these mental states have the right inconsistency properties with one another. Of course, if their view is *true*, then these states must have the right inconsistency properties, because it is inconsistent to think that

murdering is wrong and to think that murdering is not wrong. But that is just to say that this is a constraint of adequacy on their view, not to say that they are able to explain it. Cognitivists, on the other time, have the easiest of times explaining why these thoughts are inconsistent. They are inconsistent because they are beliefs toward inconsistent contents, and belief is inconsistency-transmitting.

Horgan and Timmons say that the states that they postulate are 'logically complex', but that isn't really right. What is complex, is Horgan and Timmons' syntax for designating these states. Each state must also play a certain role, that can be specified in terms of other, simpler, states. For example, the state corresponding to **¬O[ ]** must be inconsistent with the state corresponding to **O[ ]** when borne to the same content. That role is complex, and corresponds to the complexity in the syntax. So what their view gives us, is a compositional way of generating complex definite descriptions designed to pick out the attitudes expressed by complex sentences—if there are any such attitudes.

But Horgan and Timmons give us no reason other than sheer optimism to believe that these definite descriptions refer. To take the easiest case, they give us no reason to think that there really is a distinct state, **¬O[ ]**, which has these inferential properties. So there is really nothing different between their view and the one that posits both disapproval and tolerance and takes as unexplained their *B*-type inconsistency, except that Horgan and Timmons go on to draw the inevitable conclusion that you have to do this over and over again, for every other complex construction. Their 'logic' constitutes an elegant list of the things that an expressivist view needs to explain, not an explanation of them. This is *not* an idiosyncratic feature of Horgan and Timmons' view, however. The same goes for the other expressivist accounts of 'not' I'll consider in the next two sections.

3.2 Gibbard

Unlike Horgan and Timmons, Gibbard (2003) clearly emphasizes that there is something here for expressivists to explain, rather than simply to stipulate. Like Horgan and Timmons' account, Gibbard's account provides a compositional formalism that allows us to construct complex descriptions of the state of mind that is expressed by complex sentences. But like their view, this formalism does not guarantee that there is anything satisfying these descriptions. But Gibbard's best explanation fares even worse than Horgan and Timmons'. Even though Gibbard helps himself to *B*-type inconsistency as a primitive, his account still fails to distinguish n2 from n3.

Gibbard starts by observing that no matter what it is to think that murdering is wrong, this much is certain true of it: all and only the people who think that murdering is wrong are in a state of mind that is consistent with the states of mind only of other people who think that murdering is wrong.

In particular, Gibbard imagines a fictional class of 'hyperdecided' thinkers, who have views about every possible question—yea or nay—and notices that all and only the people who think that murdering is wrong are in a state of mind that is consistent with the state of mind only of *hyperdecided thinkers* who think that murdering is wrong. So consider the set of hyperdecided thinkers—Gibbard calls them *hyperplanners*—who think that murdering is wrong. Call it *S*. All and only the people who think that murdering is wrong are in a state of mind that is consistent only with the hyperplanners in *S*.

So Gibbard proposes to use *S* as a proxy for a complex description which picks out the property of thinking that murder is wrong. To say that someone is in the mental state represented by a set, *S*, is simply to say that they are in some mental state or other, such that their state of mind is consistent only with the hyperplanners in *S*. The state expressed by 'murdering is wrong', then, is represented in Gibbard's semantics by a set of hyperplanners. To represent this state by such a set, is simply to say that it is a state—who knows what—that is consistent only with the hyperplanners in the set.

He then offers the following account of negation: it should turn out that the mental state expressed by 'murdering is not wrong' should be minimally inconsistent with ('disagree' with, in Gibbard's sense) the state of mind expressed by 'murdering is wrong'. So, Gibbard says, if 'murdering is wrong' is associated with some set *S*, then 'murdering is not wrong' should be associated with the set of hyperplanners who are *not* in *S*. To unpack this, it is simply to say that 'murdering is not wrong' expresses a state of mind that is consistent only with hyperplanners whose state of mind is *not* consistent with the state of mind expressed by 'murdering is wrong'. *Ipso facto*, 'murdering is not wrong' expresses a state of mind that is inconsistent with the state of mind expressed by 'murdering is wrong'.

But this is not an *explanation*. By assigning the complement set of hyperplanners to a negated sentence, all that Gibbard's account does, is to stipulate that it is to express a state of mind that is inconsistent with the state of mind expressed by the original sentence. But it does nothing to tell us what that state of mind is like, or *why* it is inconsistent with the state of mind expressed by the original sentence. Gibbard's formalism gives us a way of generating complex definite descriptions in order to pick out the states of mind expressed by complex sentences, but no grounds to think that those descriptions actually refer, other than sheer optimism.

In fact, Gibbard's account has an even worse problem. His problem arises because he assumes that hyperplanners are always decided either to do *A* or to not do *A*, for any action *A*. So a hyperplanner never thinks that murdering is not the thing to do, without thinking that not murdering is the thing to do. They are never neutral between any options. This means that the set of hyperplanners with whom you disagree when you think that murdering is not wrong is the same as the set of hyperplanners with whom you disagree when you think that not murdering is wrong. And so despite helping himself to

everything that it looks like expressivists need to explain, Gibbard's account still fails the test of the negation problem. It still fails to distinguish n2 from n3.[6]

To be clear, my objection to Gibbard does not depend on this last point. My objection is that *even if his account worked formally*, which it does not, it leaves unexplained why 'murdering is wrong' and 'murdering is not wrong' are inconsistent, because it leaves unexplained why the underlying attitudes expressed by these sentences disagree with one another. His account stipulates that 'murdering is not wrong' is to express a state that *is* inconsistent with the state expressed by 'murdering is wrong', but like Horgan and Timmons, he has only optimism to offer, in favor of the hypothesis that there really is such an attitude.

3.3 Dreier

The best extant expressivist solution to the negation problem is due to Dreier (2006). Dreier's solution is supposed to be a fix to Gibbard's. What Dreier proposes, is to take as primitive a distinction between *indifference* and *undecidedness*. Hyperplanners, Dreier suggests, can be indifferent, even though they can't be undecided. Indifference is not a matter of having failed to make up your mind about what to do (which hyperplanners never fail to do, by definition). It is a matter of having made up your mind that it doesn't matter.

If hyperplanners can be indifferent without being undecided, Dreier argues, then permissibility without obligation, for hyperplanners, corresponds to indifference among the highest-ranked options. By allowing for hyperplanners who think that murdering is not wrong (who rank it as *one* of their top options) without thinking that not murdering is wrong (because they are indifferent between murdering and not murdering, ranking both as top options), Dreier allows for a difference in which hyperplanners you can disagree with if you think that murdering is not wrong, compared to if you think that not murdering is wrong. So Gibbard's account would go through as before, without conflating n2 with n3.

The problem that Dreier himself notes is that it may turn out that the distinction between indifference and indecision that he needs to appeal to is, like *B*-type inconsistency, something that expressivists need to *explain*, rather than something to which they have a right to appeal. But ignore that. Even if this distinction is one that makes perfect sense on expressivist grounds, this solution still helps itself to everything that expressivists really need to explain, because following Gibbard's account, it helps itself to *B*-type inconsistency.

To see why, suppose that you think that murdering is wrong. This state is represented by the set of hyperplanners who it does not disagree with. Intuitively, this should be the set of hyperplanners who think that murdering is wrong. But given Dreier's picture, there are three relevant sets of hyperplanners. There are those who think that murdering is wrong, those who think that not murdering is wrong, and those who are indifferent between

murdering and not murdering. In order for the set of hyperplanners to correctly represent your state, its members must disagree with the hyperplanners who are indifferent as well as those who think not murdering is wrong. But that means that this disagreement can't be mere *A*-type inconsistency. It has to include the *B*-type inconsistency that holds between disapproval of and tolerance of murder.[7]

What is the moral, here? The problem on which I have been focusing is not simply that expressivists have typically had a hard time distinguishing Unwin's n1, n2, and n3, although as we saw with Gibbard, they have had that. The problem is that existing expressivist views have been unable to tell us *why* normative sentences are inconsistent with their negations, because they have been unable to tell us why the attitudes they express are inconsistent. In fact, the accounts of Horgan and Timmons, Gibbard, and Dreier don't even tell us what state *is* expressed by 'murdering is not wrong', except to tell us, effectively, that it is the state—whatever it is—that is inconsistent with the state expressed by 'murdering is wrong'. This is *not* a semantic theory; it is merely a list of what expressivists would like from a semantic theory.

4.1 The Basic Expressivist Maneuver

I think that none of these looks remotely satisfactory as an expressivist explanation of why 'murdering is wrong' and 'murdering is not wrong' are inconsistent. None answers the basic question of what makes disapproval and tolerance of murdering inconsistent with one another. Each posits *that* there are such mental states that are inconsistent with one another, but none explains *why*. Fortunately for expressivists, however, it is possible to give a rather elegant solution to the negation problem that explains everything that we want to explain, appeals only to expressivist-respectable materials, and generalizes to solve other and bigger problems for expressivism.

To see the general shape of the strategy that is required, all that we need to do is to employ the *basic expressivist maneuver*. The basic expressivist maneuver is simple. Whenever you are encountered with a problem, what you do is to ask yourself what it would take to reconstruct the same problem for ordinary descriptive language. Since there is obviously no such problem for ordinary descriptive language, you use this in order to isolate the feature of your view that is creating the problem. And then you construct an answer to the problem that is based on your understanding of why ordinary descriptive language avoids the problem. This is the same kind of procedure that expressivists followed in response to the original embedding problem posed by Geach and Searle (in section 1.1). They pointed out that 'grass is green' and 'grass is not green' have distinct truth-conditions, but that that doesn't mean that their words mean different things. And then they tried to employ this lesson in explaining why expressivists don't face such a problem, either. It is also the same procedure that motivates expressivists' distinction

between 'expressing' and 'reporting' which enables them to avoid the biggest problems for cognitivist speaker subjectivism.[8]

It turns out to be easy to see how to reconstruct a negation problem for ordinary descriptive language. The key assumption that has always been made by expressivist theorists is that corresponding to each normative predicate there must be some distinct non-cognitive attitude, such that the sentence ascribing that predicate to some subject expresses that attitude toward that subject. So corresponding to 'wrong' there is disapproval; sentences of the form, 'a is wrong' express disapproval of the referent of 'a'.

Suppose that we made an analogous assumption about descriptive predicates. For each descriptive predicate, we suppose that there is some distinct (but cognitive) attitude, such that the sentence ascribing that predicate to some subject expresses that attitude toward that subject. For example, since 'grass is green' expresses the belief that grass is green, the attitude for 'green' would be, believes-green. Any sentence, 'a is green' would express the believes-green attitude toward the referent of a. Now if the believes-green attitude is really distinct and unanalyzable, what would happen? Well, we'd have a problem negating it. Compare:

g Jon thinks that grass is green.
n1 Jon does not think that grass is green.
n2 Jon thinks that grass is not green.
g∗ Jon believes-green grass.
n1∗ Jon does not believe-green grass.
n2∗ ???

(Notice that there is no n3 because 'green' is not the right kind of predicate to take a subject that admits of negations.) If believes-green were an unanalyzable attitude, then there would be no place to put the necessary negation in n2. We would have to posit a distinct and unanalyzable attitude, believes-not-green, and would no longer have any explanation of why it is inconsistent to believe-green and believe-not-green the same thing. If we took the inconsistency of these two attitudes as primitive, then we would be in the same position as the expressivists discussed in part 3.

Obviously, however, there is no negation problem with 'grass is green'. So expressivists could solve their negation problem, too, if they could only learn from that case. What makes the difference for 'green' is that the attitude that corresponds to the descriptive predicate, 'green', is not an unanalyzable attitude. Rather, it consists of a more general attitude, *belief*, together with a property, *green*. Bearing the believes-green attitude toward something is believing that it is green. So expressivists should say that disapproval is not a single, unanalyzable attitude, either. They should factor it into a more general attitude, together with a property or relation.

This should not be a surprise. What Unwin's distinction between n1, n2, and n3 showed, and what my distinction between &1, &2, and &3 emphasized, was that the expressivist accounts we were considering did not have enough *structure*. So if the problem arises from a lack of structure, there is only one solution. It is to introduce more structure. That is my solution.

4.2 Being For

Of course, the last thing that expressivists want to do is to analyze disapproval as believing wrong. So what they will have to do instead, is to analyze it in terms of a more general *non-cognitive* attitude and a descriptive property or relation. It doesn't really matter how things go, from here, but just to make things concrete, let's work with the attitude of *being for*. Being for something is bearing a very general positive attitude toward it, we can say, and we can add that if someone is for something, then they will tend to do it, other things being equal. As I will treat being for, it is therefore a *positive* attitude, rather than a negative one, and it takes *properties*—which is how I think of actions—for its object, rather than propositions. I need to include details like this in order to illustrate my positive proposal, but they are really immaterial. The important move that I am suggesting is merely to locate the right kind of *structure* in the attitudes expressed by atomic normative sentences. Everything I do in this paper can be done with a basic attitude that is negative, or one that takes propositions or states of affairs for its object, rather than properties.

Disapproval of, we can say, inspired by Gibbard (1990), is being for blaming for. (Alternatively, we could say that it is being for avoiding, or any number of other things, but we only need one example to see how the view works.) So disapproval of murdering is being for blaming for murder. But 'disapproval' was just our stipulative term for the attitude that is expressed by 'wrong' sentences. So what this is really saying is that 'murdering is wrong' expresses being for blaming for murder.

It also yields an obvious story about tolerance. Tolerating, on this view, is being for not blaming for. Someone who thinks that murdering is not wrong (who tolerates murder, in the stipulative but perhaps not the colloquial sense) is for not blaming for murder. Notice that this is not just any old story about tolerance! Together, this story about disapproval and this story about tolerance reduce their *B*-type inconsistency to *A*-type inconsistency in being for. Though they appear to be different attitudes toward the same content, murder, in actual fact they are really instances of the same attitude, being for, toward inconsistent contents—blaming for murdering and not blaming for murdering.

So if being for is an inconsistency-transmitting attitude—the kind of thing which expressivists are entitled to appeal to in their explanations—then that would suffice to explain why 'murdering is wrong' and 'murdering

is not wrong' are inconsistent. That would solve the negation problem. It would constitute an affirmative answer to Bob Hale's (1993) question, 'Can There Be a Logic of Attitudes?' The answer is, 'yes, provided that those attitudes have the right kind of structure'.

How did this work? Didn't we have an argument in part 2 that tolerance and disapproval had to be logically unrelated? No; actually all that we had was an argument that disapproval and tolerance can't be interdefined using negation. Nothing showed that they can't both be defined in terms of some third attitude, and that is what I have done. Similarly, if we went in for such things, there would be no way of interdefining the attitude of believing-green and the attitude of believing-not-green, if believing-green were unanalyzable. But there is no mystery about how to solve this problem: since believing-green is just believing to be green, we can analyze both in terms of belief.

There should be nothing surprising about this solution. All that it does, is to look for the right structural feature that allows descriptive language to avoid the negation problem, and impose that structure on normative thought. The only difference is that in order to be consistent with expressivism, we simply have to hold that the more general attitude is not belief, but a practical, non-cognitive attitude. I am calling this attitude *being for*, but structurally speaking, it doesn't matter what this attitude is, so long as normative predicates all correspond to this attitude plus some further property or relation.

This solution works by creating an extra place in which to insert the negation needed in n2, and does so *in the very same way* as this works for descriptive sentences. Compare:

w† Jon is for blaming for murdering.
n1† Jon is not for blaming for murdering.
n2† Jon is for not blaming for murdering.
n3† Jon is for blaming for not murdering.

Factoring disapproval into the more general attitude of being for plus the relation of blaming for creates an extra place in which we can insert and interpret a negation, just as factoring *believes-green* into belief and green leaves an extra place in which we can insert and interpret a negation. If the problem arises because the expressivist account has insufficient structure, there is *only one* solution: to give the expressivist account sufficient structure. And that is my solution. *Ipso facto*, my solution is the only one that works.

So this solves the negation problem for expressivism. By finding the right kind of structure in the attitudes expressed by atomic normative sentences, we can deal with them in the same way that we would deal with descriptive sentences, we can reduce the problematic *B*-type inconsistency between disapproval and tolerance to relatively unproblematic, because explicable,

A-type inconsistency in being for, and we can explain the inconsistency of non-cognitive attitudes in terms of the inconsistency of their contents, in the same way as we do for beliefs.

This solution is far more powerful than the expressivist accounts currently on the table. It explains more, by appeal to less, than any other expressivist account. And it does so in the only obvious way—by seeing why there is no negation problem for descriptive sentences, and accounting for normative sentences in a way with the same structural features. As I have noted, there will of course be several ways of implementing this basic idea, depending on how we characterize the basic non-cognitive attitude of being for, and depending on which descriptive relation we associate with each normative predicate. But any satisfactory expressivist account is going to have to have this structure. If the problem arises from a lack of structure—and it does—there is only one solution—to add structure—and this is it.

But more—I can also show that there are independent grounds to think that this is the right track for expressivism. For now that we understand how to explain inconsistencies of the form, {P,∼P}, and now that we know what we need, in order to have the right structure in order to introduce logical connectives, we now have the tools to solve bigger problems for expressivism. So to illustrate just how powerful this strategy is, I'll now show how it leads to a totally straightforward account of how to construct a compositional semantics for a simple normative language under an expressivist interpretation that allows for an explanatory expressivist account of the logical notions of inconsistency, entailment, and validity.

5.1 Generalizing

It is no mystery how my account solves the negation problem. By factoring disapproval into a more general attitude and a relation, it leaves an extra place for the negation—it goes inside the more general attitude of being for, but outside the relation. So far, this only tells us how to negate atomic normative sentences. But it is easy to see that this leads to a natural, perfectly general account of negation.

To see how, again compare ordinary descriptive language. If 'P' expresses the belief that *p*, then '∼P' should express the belief that ∼*p*, regardless of whether 'P' is simple or complex. Our account so far tells us that if 'N' is an atomic normative sentence expressing being for *n*, then '∼N' expresses being for ∼*n*. That is, the negation goes inside the state of being for, but outside everything else. But modeling on the case of belief, there is no reason why this account shouldn't be perfectly general. This leads to the following rule:

∼N: For any normative sentence 'N', if 'N' expresses being for *n*, then '∼N' expresses being for ∼*n*.

This rule will work, as long as all normative sentences turn out to express states of being for something. And so now we have an expressivist-respectable explanation of why it is that any normative sentence is inconsistent with its negation. Whatever the normative sentence is, it expresses some state of being for. And so its negation expresses a state of being for the negation of what it does. Since the contents of these states of being for are ordinary descriptive contents, ordinary descriptive logic guarantees that they are inconsistent. But then so long as being for is an inconsistency-transmitting attitude, these states are inconsistent. And so it follows that the sentences which express them are inconsistent, by the treatment of inconsistency of sentences set out in section 1.2.

Moreover, whereas all previous expressivist accounts of inconsistency simply posit that there is a kind of mental state that has the right properties, but don't actually tell us what it is, except in terms of the properties that it needs to turn out to have, this account is actually *constructive*. Other expressivist accounts don't answer the embedding problem, because they merely list the properties that the attitudes expressed by complex sentences need to have. But this account actually gives us a constructive account of what these mental states are, and uses a single general property of a single underlying kind of attitude—that of being for—in order to explain why they have the requisite properties.

Of course, that's great for explaining the inconsistency of sentences and their negations. But we ultimately want to explain the inconsistency of arbitrary sentences. How are we to do that? Well first, obviously, we need a constructive account of which attitude is expressed by arbitrary sentences. And fortunately, our account shows us how to do exactly that. Recall from section 2.2 that negation was not the only construction that we did not have enough places for, on the assumption that disapproval was unanalyzable. There was also a problem with conjunction.

In fact, there is a problem with every complex-sentence-forming construction. To see why, simply compare how such constructions work for ordinary descriptive sentences. If 'P' expresses the belief that p and 'Q' expresses the belief that q, then 'P&Q' should express the belief that $p\&q$ and 'P∨Q' should express the belief that $p \vee q$. In each case, the connective gets located inside the general attitude, but outside whatever property is associated with the descriptive predicate. The analogous rules for normative language would have to look like this:

&N: If 'N' is a normative sentence expressing being for n and 'O' is a normative sentence expressing being for o, then 'N&O' expresses being for $n\&o$.

∨N: If 'N' is a normative sentence expressing being for n and 'O' is a normative sentence expressing being for o, then 'N∨O' expresses being for $n \vee o$.

What these rules say is simple and intuitive. If thinking that murdering is wrong is being for blaming for murdering and thinking that stealing is wrong is being for blaming for stealing, then &N says that thinking that stealing is wrong and murdering is wrong is being for blaming for murdering and blaming for stealing. Similarly, ∨N says that thinking that murdering is wrong or stealing is wrong is being for blaming for murdering or blaming for stealing.

We could also introduce the material conditional and biconditional in this way (though they are harder to say informally, using gerunds), and it is relatively straightforward to generalize and include quantifiers.[9] But this is enough to illustrate the power of our approach.

5.2 Inconsistency

What we now have is the first constructive compositional expressivist semantics, for a rudimentary expressivist language with atomic normative sentences. Gibbard and Horgan and Timmons, of course, gave us compositional formalisms. But their accounts were not constructive. They merely postulated, for each compositional rule, an indefinite number of distinct attitude kinds, and assumed them to have the required properties. But this account actually tells us what mental states are expressed by normative sentences of arbitrary complexity, given our simple range of connectives. In fact, it induces a simple rule: the structure of any complex normative sentence 'N' is isomorphic to the structure of the content of the state of being for that it expresses. This rule is no mystery; it is exactly how we expect things to work for beliefs. We expect the content of a belief expressed by an ordinary descriptive sentence to be isomorphic to the logical structure of that sentence. Since we have analogous rules for normative sentences, we get the same thing for normative sentences. It is straightforward to verify this by a trivial induction on sentence complexity, using our three compositional rules, ∼N, &N, and ∨N.

This means that inconsistency for normative sentences is easy to explain. We say that two sentences are inconsistent just in case the attitudes that they express are. Since normative sentences all express states of being for, it follows that two normative sentences are inconsistent just in case the states of being for that they express are. But assuming that being for is an inconsistency-transmitting attitude, this reduces to the inconsistency of the contents of those states of being for. But our compositional semantics induces an isomorphism between the contents of the states of being for expressed by sentences and those sentences themselves. So any two sentences that we would want to turn out to be inconsistent, because they would if governed by ordinary descriptive logic, will express states of being for whose contents have that very same structure and *are* governed by ordinary descriptive logic. So that will guarantee that the sentences are inconsistent.

That seems a bit abstract. So let's try an example. Take the sentences, 'murdering is wrong' and 'murdering is not wrong and stealing is better than murdering'. We have said that 'murdering is wrong' expresses being for blaming for murdering; let us say that 'stealing is better than murdering' expresses being for preferring stealing to murdering. So by our compositional rule ∼N, 'murdering is not wrong' expresses being for not blaming for murdering, and by rule &N, 'murdering is not wrong and stealing is better than murdering' expresses being for not blaming for murdering and preferring stealing to murdering. So the mental states expressed by these two sentences are states of being for blaming for murdering, and of being for not blaming for murdering and preferring stealing to murdering. But blaming for murdering is inconsistent with not blaming for murdering and preferring stealing to murdering, by ordinary descriptive logic. So assuming that being for is inconsistency-transmitting, these states are inconsistent. And so it follows that the sentences are, as well.

That gives us an account of inconsistency, one of whose very nice features is that it reduces the inconsistency of normative sentences to ordinary descriptive inconsistency in the contents of the states of being for that they express. But it falls short of an account of logical inconsistency, because logical inconsistency should be inconsistency guaranteed by form, but so far our account allows for inconsistencies that are guaranteed only by the intended interpretation of our normative predicates and terms. For example, as it stands, our account explains why 'defenestration is wrong' and 'murdering someone by pushing them out of a window is not wrong' are inconsistent. This is right, of course, but this is not logical inconsistency. It is due only to the fact that on the intended interpretation, 'defenestration' picks out the same action as 'murdering someone by pushing them out of a window'. An account of logical inconsistency should avoid this.

So we can say that arbitrary normative sentences *A* and *B* are *logically* inconsistent just in case they are inconsistent under any consistent interpretation of which state of being for is expressed by their atomic normative sentences. It is a simple exercise to verify that this yields the right results. Of course, ultimately we want an account of inconsistency that will apply to arbitrary sets of *n* sentences, not merely an account of pairwise inconsistency. But that is easy to obtain, as well. We can say that set $\{A_1,\ldots A_n\}$ is logically inconsistent just in case the set $\{A_1\&\ldots\&A_{n-1},A_n\}$ is pairwise logically inconsistent.

And then, assuming that nothing pushes us to deny excluded middle for our simple normative language, we can define logical entailment and logical validity in the obvious ways. We say that A logically entails B just in case {A,B} is a logically inconsistent set, and that the argument, $P_1,\ldots P_n;C$ is logically valid just in case $\{P_1,\ldots P_n,\sim C\}$ is a logically inconsistent set. In this way, we can easily generalize our solution to the negation problem to

an account of how to do logic for our simple normative language under an expressivist semantics.

5.3 Modus Ponens

Just as an example, we can observe how this works with Geach's original *modus ponens* argument, which was:

P1 Lying is wrong.
P2 If lying is wrong, then getting your little brother to lie is wrong.
C Getting your little brother to lie is wrong.

What we want to show is not merely why this argument is 'truth-preserving', but why it is logically valid. By our account, it is logically valid just in case the set {P1,P2,∼C} is a logically inconsistent set, which by our definition, is true just in case {P1&P2,∼C} is pairwise logically inconsistent.

Logical inconsistency, recall, is indifferent to interpretation. So without loss of generality, let us assume that P1 expresses a state of being for α and C expresses a state of being for β. So by rule ∼N, ∼C expresses being for $\sim\beta$. Similarly, 'lying is not wrong' expresses being for $\sim\alpha$. Let's assume that the conditional is a material conditional, and hence that P2 can be rewritten, 'getting your little brother to lie is wrong or lying is not wrong'. Then it follows from rule ∨N that P2 expresses being for $\beta\vee\sim\alpha$. And so by rule &N, it follows that P1&P2 expresses being for $\alpha\&(\beta\vee\sim\alpha)$. Since we are assuming that being for is inconsistency-transmitting, it follows that P1&P2 is inconsistent with ∼C under this arbitrary interpretation just in case $\sim\beta$ is inconsistent with $\alpha\&(\beta\vee\sim\alpha)$. But these are governed by ordinary descriptive logic, so it is easy to see that they are inconsistent. So since the choice of α and β was arbitrary, it follows that {P1&P2,∼C} is inconsistent under any interpretation, and hence logically inconsistent.

And so our semantics can explain why Geach's argument is logically valid, given only one assumption: that being for is an inconsistency-transmitting attitude. But that is the kind of thing that I argued in section 1.3 that expressivists should be entitled to appeal to. At least, it is a more respectable thing for them to appeal to than brute *B*-type inconsistency, because there are perfectly respectable models of inconsistency-transmitting attitudes, in belief and intention. Expressivists therefore need only that being for turns out to have this broad feature of belief and intention, in order to be able to account for logical validity in a normative language under an expressivist interpretation. This *has* to be the way for expressivists to go.

6.1 Further Lessons

I think that the program that I have just suggested in outline has to be the most promising approach for expressivists. Previous expressivist views, I've argued, have not even been able to explain the inconsistency of 'murdering is wrong' and 'murdering is not wrong'. They all take *B*-type inconsistency for granted, and merely stipulate that there must be some further attitude, tolerance, that has the right inconsistency properties. They tell us nothing about what tolerance *is* that could help to explain *why* it has these properties, other than to say that it is the mental state that has them. My account, on the other hand, is constructive. For any complex normative sentence, the account that I've offered can tell us exactly what mental state is expressed by that sentence. Each one is a state of being for, and the compositional semantics tells us *what* it is for.

This yielded a constructive compositional semantics that was exactly as powerful as the analogous story might have gone about which attitudes are expressed by ordinary descriptive sentences. All such states are beliefs, and the compositional semantics merely tells us what the *contents* of those beliefs are. It is as a result of this feature that we were able to get a general account of inconsistency and hence of logical inconsistency and logical validity for our simple, purely normative, expressivist language. Since all of the sentences in our simple normative language express states of being for, that made it easy to account for the inconsistency of sentences by appealing to a single basic property of the attitude of being for—that it is inconsistency-transmitting. This was analogous to the explanation of the inconsistency of beliefs in terms of their contents, by appeal to the assumption that belief is inconsistency-transmitting. It is the feature that made the account work.

All of this would be fantastic if we spoke one normative language and one descriptive language, each of which had its own complex sentences and was subject to its own logical inconsistency relations. But the salient problem for expressivists is that we speak a language with both normative and descriptive predicates, which can combine under a single set of logical connectives to make complex sentences with both descriptive and normative parts. What are we to say about such sentences? What kind of attitude do they express?

The problem is that if they express beliefs, then we lose any prospect of generalizing our explanation of inconsistency in order to explain how they can be inconsistent with purely normative sentences, which express states of being for. On the other hand, if they express states of being for, then we lose any prospect of generalizing our explanation of inconsistency in order to explain how they can be inconsistent with ordinary descriptive sentences. Yet 'murdering is wrong and grass is green' is inconsistent both with 'murdering is not wrong' and with 'grass is not green'. And if we say that complex normative-descriptive sentences express some third kind of attitude, neither

belief nor being for, then we make no progress at all, starting with our original problems all over again.

6.2 Advertisement

So it seems clear that the only way to apply the advantages of the account that I've sketched here, on which we can reduce the explanation of the inconsistency of arbitrary sentences to the inconsistency of the contents of the attitudes that they express, is to allow that all sentences express the same general kind of attitude. For cognitivists, this is easy to do, for the attitude is simply belief. But for expressivists, this turns out to be tricky. For descriptive sentences, according to expressivists, express beliefs. So if they are to express states of being for, then it must turn out that beliefs are really states of being for – that belief needs to be analyzed in terms of a non-cognitive attitude.

This would be a surprising result! Expressivism, after all, is often motivated by a motivational argument, which posits a deep divide between belief and desire, holding that belief is the wrong kind of state to have a capacity to motivate. But now, I am arguing, the most promising approach for expressivists to the embedding problem places on them some very strong pressure to try to analyze belief in terms of a non-cognitive attitude! It is important to distinguish this conclusion from the more general one, that supplying a uniform account of the sentential connectives commits expressivists to giving a non-standard semantics even for complex ordinary descriptive sentences.[10] The view I am suggesting here is a much more radical one. It is that if the foregoing arguments are correct, then expressivists are committed to a radical view not only about *semantics*, but in the *philosophy of mind*—the view that belief itself needs to be analyzed in terms of a further noncognitive attitude.

Nevertheless, surprising though this result may be, I think that such an account can be given, and the preceding constitutes an argument that expressivists should give it. In related work,[11] I consider how to develop such an account, and consider one which yields a constructive compositional expressivist semantics for a language with both normative and descriptive predicates and the expressive power of the predicate calculus. For such a language, we can account for logical inconsistency in a generalization of the same way done here, and I claim that this is the most satisfactory way to carry out the research program of Hare, Blackburn, and Gibbard.

Even then, of course, the account raises further complications, and although I think it is what expressivists *should* say, I don't think that it is true. I think, in fact, that the more we discover about the commitments of expressivism, the more implausible it seems, as an account of how normative terms work in natural languages. But that is just to clarify that I am not, myself, an expressivist. The important points in this paper show not that expressivism might be true, but which problems don't demonstrate it to be false. For now

I claim merely to have shown how expressivists *can* solve their problem with negation, and to have argued that by all existing evidence, this is how they *should* do so.[12]

Notes

[1] There is an important complication, here, due to the fact that on many expressivist views, expression is a speech act, and so the parts of a complex sentence don't really express the attitudes that they do when unembedded, which means that the attitude expressed by the complex sentence really can't be, strictly speaking, a function of the attitudes expressed by its parts (for they don't express any). I'm going to ignore this complication here; anyway, I've shown how expressivists can avoid it, by appealing to the right account of the *expression* relation, in 'Expression for Expressivists'.

[2] See especially Hare (1970), Blackburn (1984), Schueler (1988), Blackburn (1988), Gibbard (1990), Hale (1993), van Roojen (1996), Unwin (1999), (2001), Sinnott-Armstrong (2000), Gibbard (2003), Dreier (2006), and Schroeder (2008 c).

[3] One might hope that it is also sufficient, but it turns out that some expressivist accounts explain why $\{P,P\supset Q,\sim Q\}$ is in some sense inconsistent in ways that leave unexplained or even in doubt why someone who accepts the premises of a *modus ponens* argument is committed to accepting its conclusion. For example, this is a fault in some 'higher-order attitudes' accounts, as I discuss in Schroeder (forthcoming). And it is also a fault in some 'hybrid' expressivist accounts such as that of Ridge (2006), as is discussed in van Roojen (2005).

[4] See, for example, Harman (1976), (1986), Velleman (1989), Wallace (2001), Broome (forthcoming), Setiya (2007). See also Bratman (forthcoming a) and (forthcoming b). Some of these theorists claim to reduce other kinds of coherence requirements on intention to requirements on belief, rather than the consistency requirement.

[5] For discussion of how even these apparent *B*-type inconsistencies between attitudes might reduce to *A*-type inconsistencies, see Schroeder (2008 b), chapter 7.

[6] See Dreier (2006) for further discussion of this last point.

[7] It's also worth noting that Gibbard can't accept Dreier's friendly fix, since he needs the assumption that hyperplanners can't be indifferent in order for his argument that normative terms pick out natural properties to go through. This argument and its consequences are one of the main contributions of Gibbard (2003).

[8] See Schroeder (2008 a) for further discussion.

[9] I show how to do so, and present the semantics more rigorously, in Schroeder (2008 b).

[10] See Hale (1993), Kölbel (2002), and Schroeder (2008 a) for discussion of this point.

[11] Schroeder (2008 b).

[12] Special thanks to Jeff Horty, Jamie Dreier, Jacob Ross, Stephen Finlay, Jan Gertken, Scott Soames, Sari Kisilevsky, Paul Pietroski, Barry Lam, Mike McGlone, Matt King, Ryan Wasserman, an anonymous referee for *Noûs*, and to the members of my spring 2006 graduate seminar on expressivism at the University of Maryland College Park.

References

Blackburn, Simon (1973). 'Moral Realism.' Reprinted in Blackburn (1993).

—— (1984). *Spreading the Word*. Oxford: Oxford University Press.

—— (1988). 'Attitudes and Contents.' *Ethics* 98(3): 501–517.

—— (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.

—— (1998). *Ruling Passions*. Oxford: Oxford University Press.

Bratman, Michael (1993). 'Cognitivism about Instrumental Reason.' Reprinted in Bratman, *Faces of Intention*. Cambridge: Cambridge University Press (1999), 250–264.

—— (forthcoming a). 'Intention, Belief, Theoretical, Practical.' Forthcoming in Timmerman, Skorupski, and Robertson, eds., *Spheres of Reason*.
—— (forthcoming b). 'Intention, Belief, and Instrumental Rationality.' Forthcoming in David Sobel and Stephen Wall, eds., *Reasons for Action*.
Broome, John (forthcoming). 'The Unity of Reasoning?' Forthcoming in Timmerman, Skorupski, and Robertson, eds., *Spheres of Reason*.
Dreier, James (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.
Geach, P.T. (1960). 'Ascriptivism.' *Philosophical Review* 69: 221–225.
—— (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.
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.
Harman, Gilbert (1976). 'Practical Reasoning.' Reprinted in Harman, *Reasoning, Meaning and Mind*. Oxford: Oxford University Press (1999), 46–74.
—— (1986). *Change in View*. Cambridge: MIT Press.
Horgan, Terry, and Mark Timmons (2006). 'Cognitivist Expressivism.' In Horgan and Timmons, eds., *Metaethics after Moore*. Oxford: Oxford University Press, 255–298.
Kölbel, Max (2002). *Truth Without Objectivity*. New York: Routledge.
Ridge, Michael (2006). 'Ecumenical Expressivism: Finessing Frege.' *Ethics* 116(2): 302–336.
Schroeder, Mark (2008a). 'Expression for Expressivists.' *Philosophy and Phenomenological Research* 76(1): 86–116.
—— (2008b). *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.
—— (2008c). 'What is the Frege-Geach Problem?' *Philosophy Compass* 3/4: 703–720.
—— (forthcoming). *Noncognitivism in Ethics*. Book manuscript. Under contract with Routledge.
Schueler, G.F. (1988). 'Modus Ponens and Moral Realism.' *Ethics* 98(3): 492–500.
Searle, John (1962). 'Meaning and Speech Acts.' *Philosophical Review* 71: 423–432.
Setiya, Kieran (2007). 'Cognitivism about Instrumental Reason.' *Ethics* 117(4): 649–673.
Sinnott-Armstrong, Walter (2000). 'Expressivism and Embedding.' *Philosophy and Phenomenological Research* 61(3): 677–693.
Unwin, Nicholas (1999). '*Quasi*-Realism, Negation and the Frege-Geach Problem.' *The Philosophical Quarterly* 49(196): 337–352.
—— (2001). 'Norms and Negation: A Problem for Gibbard's Logic.' *The Philosophical Quarterly* 51(202): 60–75.
van Roojen, Mark (1996). 'Expressivism and Irrationality.' *The Philosophical Review* 105(3): 311–335.
—— (2005). 'Expressivism, Supervenience, and Logic.' *Ratio* 18(2): 190–205.
Velleman, David (1989). *Practical Reflection*. Princeton: Princeton University Press.
Wallace, R. Jay (2001). 'Normativity, Commitment, and Instrumental Reason.' *Philosophers' Imprint* 1, no. 3.