

Expression and Guidance in Schroeder's Expressivist Semantics

Note: this is a penultimate draft of a paper forthcoming in *Erkenntnis*. Please quote or cite the published version.

Abstract

Mark Schroeder's expressivist program has made substantial progress in providing a compositional semantics for normative terms. This paper argues that it risks achieving this semantic progress at the cost of abandoning a key theoretical motivation for embracing expressivism in the first place. The problem can be summarized as a dilemma. Either Schroeder must allow that there are cases in which agents are in disagreement with one another, or can make *valid* inferences, but that these disagreements or inferences are not expressible in natural language; or his version of expressivism must abandon one of the key theoretical advantages expressivist theories seemed to possess over cognitivism, the ability to provide a very straightforward explanation of the action- and attitude-guiding role of normative judgments.

1. Introduction

In his (2008a; 2008b; and 2010a) Mark Schroeder presents an expressivist semantics which he argues can explain inconsistency and related notions such as validity and entailment, and which is rich enough to provide a compositional account for the standard operators and connectives of first-order logic, thereby solving the negation and Frege-Geach problems. The theory relies on two theoretical posits: the attitude of *being for*, and *A-type inconsistency*—which he later calls *non-Moorean inconsistency*

(following van Roojen 1996). Both notions will be explained in more detail in the following sections. For now, it is enough to say that the being-for attitude is the non-cognitive attitude expressed by normative natural language utterances; whereas A-type inconsistency is the type of rational conflict that holds between token attitudes of being for, and which explains how utterances can be semantically—*not* merely pragmatically—inconsistent with one another, and thus how normative assertions can *entail* or be *valid consequences* of other assertions. These two notions, then, jointly undergird the semantic program.

The justification for taking on these posits is their explanatory power. This paper will argue that in fact they have less explanatory power than initially seems. Either Schroeder must allow that there are cases in which agents are in disagreement with one another, or hold token instances of the being-for attitude on the basis of valid inference from other token instances, but are unable to express the disagreement or inference in natural language; or his version of expressivism must abandon one of the key theoretical advantages expressivist theories seemed to possess over cognitivism, the ability to provide a very straightforward explanation of the action-and attitude-guiding role of normative judgments, at least for an important subclass of those judgments, those that Gibbard identifies as “flavorless.”

Schroeder’s central goal in *Being For* is to develop an adequate semantic theory for expressivism—so it should be noted that both horns of the dilemma present “extra-semantic” problems. They do, nonetheless, bear on the credibility of the semantic theory. It may be, for example, that expressivist theories can only purchase semantic adequacy at the price of embracing mysteries about the limits of

our language—holding for example that there are valid inferences that cannot be made the objects of public discussion or debate. Or it may turn out that an adequate semantics is only to be had by giving up on those features of expressivism that made it attractive in the first place. In either case, the objection is that Schroeder’s version of expressivism gains explanatory power in the domain of meaning, but at the cost of acquiring new explanatory burdens elsewhere.

Schroeder’s expressivism is, to date, arguably the most comprehensive and sophisticated form on offer. (This is not to say that there are no promising alternatives, only that at present they lack the scope and systematicity.)¹ The dilemma presented here is thus potentially revealing of the pitfalls faced by expressivist theories in general. Expressivism ties *meaning* to reasoning and motivation much more directly than an orthodox representational semantics. This forces the expressivist to keep her eyes on a broader range of issues in developing her semantic theory, otherwise she risks solving the semantic problems of expressivism, but only at the cost of pushing the theoretical bump to another part of the rug, acquiring additional, potentially unwelcome commitments elsewhere—most likely in her moral psychology.

¹ See, for example, (Charlow 2014; Baker and Woods 2015; Schwartz and Hon 2015; Alwood 2016; and Shiller 2016).

Ridge’s (2014) proposal does arguably possess the scope and systematicity of Schroeder’s—but as *hybrid* expressivist theory, it is less ambitious in how significantly it departs from orthodox semantic theory, a point that Ridge offers in its favor. The earlier, quite detailed expressivist projects of Blackburn and Gibbard have also received defense in (Sinclair 2011; and Gibbard 2012: 273-92).

The paper will begin by discussing earlier expressivist accounts to explain inconsistency and negation. This will motivate Schroeder's explanatory posits, and set the stage for the problems they leave unaddressed.

2. Gibbard's Negation Problem

'Murdering is not wrong' and 'Not murdering is wrong' have distinct meanings. Thus, an adequate semantics for moral terms must be able to distinguish sentences with the following structure:

(a) $\sim F(x)$

(b) $F(\sim x)$

This section will show why Gibbard risks conflating (a) and (b) for the central normative predicates of his theory. (Subsequent sections show that he has no particular difficulty accounting for those normative terms he treats as more derivative, such as the term 'wrong'.) The central normative predicates are those that express "flavorless endorsement" (1990: 67). These are predicates that pick out an action as *the thing to do*, or an attitude as *the one to have*. They indicate that the action or attitude *makes sense*. For example, when I tell someone they *ought* to watch TV shows that they enjoy, rather than ones they find boring, I am not stating a moral obligation. I am offering normative advice of the most general kind—I am saying which choice makes sense.

This paper will primarily use ‘makes sense’ for the term of flavorless endorsement. One could also use ‘reasonable’ in the sense offered in (Scanlon 1998), following current philosophical usage, or follow Gibbard (1990) and use ‘rational’.² The paper will also, on occasion, follow Gibbard and use ‘ought’, where this term is not explicitly relativized to refer to moral or prudential obligations; this paper will treat ‘One ought to x’ as equivalent to ‘x-ing makes sense’ or ‘x-ing is reasonable’.

For Gibbard, “flavored” normative predicates, such as moral predicates, aesthetic predicates, and predicates of propriety, are to be understood in terms of ‘ought’ (1990: 51-2). According to Gibbard, ‘x is wrong’ is roughly equivalent to ‘One ought to become angry with those who x’ or ‘Becoming angry with those who x is reasonable’ (41-2).³

Given this analysis, Gibbard’s theory depends on offering an expressivist account of flavorless endorsement—expressivism for the flavored follows.

Gibbard holds that the judgment that some act or attitude makes sense expresses a plan to have the attitude or perform the act in the relevant circumstances.⁴ So on Gibbard’s semantics ‘One ought to x’ or ‘x is reasonable’ expresses:

² Gibbard is clear that ‘rational’ is meant to pick out the thing to do, the act or attitude that would make most sense.

³ Note that this formulation skips over the Wrong Kind of Reasons problem. For some discussion of this problem, see (D’Arms and Jacobson 2000; Rabinowicz and Rønnow-Rasmussen 2004; Hieronymi 2005; Schroeder 2010b; and Sharadin 2013). Issues surrounding the Wrong Kind of Reasons problem will come up later (see footnote 14), but these complications do not affect the main argument.

⁴ This skips over the formulation in terms of world-norm pairs (Gibbard 1996), and the device of hyper-decided agents (2003: 57). The simplification is not relevant to the argument at this point, which primarily aims to show the motivations for Schroeder’s theory.

PLAN (x)

The problem is that ought-statements also admit of internal and external negations, which presumably have distinct meanings. We can say, for example, ‘One ought to not x ’. Presumably this would express the following:

PLAN ($\sim x$)

But we can also say ‘It’s not the case that one ought to x ’. It’s not clear, however, what planning state this could express. Perhaps it is:

PLAN ($\sim x$)

But then we have conflated internal and external negation for flavorless ought-statements.

Could we instead make the sentence express the following?

\sim PLAN (x)

But this leads to difficulties. First, x is a semantic object, and so we have some idea of what it is to negate it. It is not clear what it is to negate an attitude, however (Dreier 2006). Perhaps we should interpret ‘It is not the case that one ought to x ’ as expressing the absence of a plan to x . But then we have no way of distinguishing a

person who definitively judges that it is not the case that one ought to x from a person who has no opinion about whether one ought to x (Unwin 2001; Dreier 2006; and Schroeder 2008a, and 2008b).⁵

Schroeder's semantics is intended to improve on Gibbard's in several respects, but the first of these is making sufficient room for distinguishing internal and external negation where Schroeder's cannot.

3. Schroeder's Solution

Schroeder's solution to the negation problem posits the existence of a master conative attitude, *being for*. This attitude takes a relation towards objects as its content. The meaning of normative predicates is then explained in terms of the relations that can be the objects of the being for attitude. 'X is better than Y' expresses being for preferring X to Y ; and 'Murder is wrong' expresses being for blaming for murder (2008b: 57-8).

It should be noted that this attitude is a theoretical posit; it is not identified with any familiar attitude from folk psychology. What we are told about it is that it takes relations towards objects as its content (at least those instances expressed in utterance), and it can stand in relations of A-type inconsistency to other instances of being for (the nature and importance of A-type inconsistency will be explained in the subsequent section). Schroeder's justification for this posit is theoretical utility. With it, he can account for the difference between 'Murder is not wrong' and 'Not

⁵ Dreier's (2006) solution to this problem is outside the scope of this paper.

murdering is wrong' as follows. The first sentence expresses the following:

FOR (\sim blaming (murder))

While the second expresses:

FOR (blaming (\sim murder))

More generally, Schroeder tells us that for any normative predicate 'F', the sentence 'F(x)' will express an attitude with the following structure:

FOR (ϕ (x))⁶

Where ϕ is some relation (again, the initial examples offered being attitude-types, although other types of relations, we will see, are possible). This allows us to distinguish (a) from (b) because we can distinguish the attitude of *FOR* ($\sim\phi$ (x)) from *FOR* (ϕ (\sim x)).

Schroeder describes the idea behind his solution thus: "If the problem arises from a lack of structure [in the attitudes], there can be only one solution: to add structure" (2008b: 61).

⁶ It would be more correct to say that it has the structure *For* (ϕ (... , x)). As Schroeder (2012) puts it: "the semantic value of each *n*-place predicate is an *n+1*-place relation." Interested readers should also see chapter 6 of *Being For*. The simplification here does not affect the main argument, however.

But recall Gibbard’s account of the meaning of ‘wrong’. “...[T]o think an act wrong is to accept norms for guilt and resentment that, prima facie, would sanction guilt and resentment if the act were performed” (1990: 47). Gibbard appears committed to building exactly the same kind of structure as Schroeder into the attitude. ‘X is wrong’, on the Gibbard account, expresses:

PLAN (anger (x))

Presumably this sort of structure will be available to Gibbard for all flavored predicates. He should, then, be able to distinguish the difference between (a) and (b) for these predicates.

Gibbard’s solution to the negation problem fails because he cannot extend this structure to the attitudes expressed by the generic ‘ought’. But Schroeder provides no account of how his own solution would cover these cases.⁷ Part of this failure seems due to the concern that ‘ought’ is a modal term rather than a predicate, which introduces its own semantic difficulties (Schroeder 2012). But there are flavorless *predicates*: ‘rational’, ‘reasonable’, ‘the thing to do’, or ‘makes sense’. The treatment needs to be extended to these.

As we will see, there is a straightforward way in which Schroeder’s account can be extended: simply identify the ϕ that stands to the flavorless predicate as *blame* does to ‘wrong’. The real problem isn’t a semantic one; rather, it becomes unclear on

⁷(Skorupski 2012) reads Schroeder’s semantics as effectively the same as Gibbard’s, with the being for attitude corresponding to ‘ought’ in the way the planning attitude corresponds for Gibbard. See (Schroeder 2012; and Köhler 2012) for arguments that this is a mistake.

Schroeder's account how judgments involving flavorless predicates are supposed to interact with judgments involving the flavored. To explain this problem, we must first look at how Schroeder solves the other half of the negation problem.

3.1. The Other Half of the Negation Problem: The Right Kind of Inconsistency

There is another half to the negation problem. Expressivists do not merely risk conflating internal and external negations. They also risk conflating pragmatic or implied contradictions with semantic or literal contradictions.

A longstanding criticism of proposed expressivist solutions to the Frege-Geach problem is that they conflate distinctive types of rational failings. The person in a state of mind such that he would affirm both 'P' and '~P' is guilty of a logical mistake. Expressivist accounts are charged, however, with making the mistake seem more akin to one of akrasia, hypocrisy, or indecisiveness (cf. Schueler 1988; Wright 1988; Zangwill 1992; and van Roojen 1996).

By itself the force of this objection may seem minor. Who cares if the expressivist treatment of accepting inconsistent sentences doesn't make the particular irrationality in question *feel* logical enough? But the objection becomes more pressing when we look at van Roojen's (1996) use of the Moore paradoxical sentence, "It's raining, but I don't believe it is raining." This sentence, notoriously, sounds as though the speaker is contradicting herself, even though there is no contradiction. Both conjuncts could well be true. This presents problems for expressivist accounts of negation on two fronts.

First, the expressivist needs to tell us what it is for one sentence to be inconsistent with, or contradict, another. Notice that they cannot explain this in terms of realist materials such as incompatible truth conditions, or the fact that the sentences describe states which are not compossible. (Expressivists may ultimately give a quasi-realist account, but to qualify as *quasi-realist*, the realistic materials cannot bear explanatory weight.) There should still be some sense, however, in which a sentence “rules out” those sentences which are inconsistent with it. Given the expressivist account of meaning in terms of psychological states, the most natural way of developing this idea is that the attitude expressed by one sentence rules out, as a psychological matter, the attitude expressed by an inconsistent sentence.⁸ The agent who had both attitudes would be at odds with herself, in disagreement with herself, would suffer a fractured psychology, or would be otherwise psychologically *inconsistent* (Gibbard 1996 and 2003; Blackburn 1998; Dreier 2006; Baker and Woods 2015; and Silk 2015).

But some failures of self-knowledge seem to involve failings of rational coherence. To believe that it is raining while simultaneously believing of yourself that you don’t believe it is raining, for example, is intuitively to have a fractured psychology, one at odds with itself (Wittgenstein 1953: Part II; Shoemaker 1996; McGeer 1996; and Moran 2001). But, van Roojen points out, this shows that the range of sentences expressing rationally conflicting attitudes is wider than the sentences which are literally inconsistent.

Second, the expressivism is often motivated by a strategy of understanding

⁸ But see (Charlow 2014; and Alwood 2016) for alternatives.

the peculiar features of the normative not in terms of special normative properties, but rather the peculiarities of normative language: we use such language not primarily to describe, but to prescribe, advise, praise, censure, permit, and forbid. Of course, a normative realist can also allow that normative sentences are used to do those things. In order to obviate any need to appeal to normative properties, the expressivist is pressured to hold that the meaning of normative utterances is explained by their communicative or conversational role, the kinds of speech acts, or in short, *use*. But if Nate tells me “It is raining,” I can *use* the sentence “I don’t believe that,” to disagree, to contradict him in the context of our conversation. This shows that the range of pragmatic or communicative inconsistencies is wider than the range of sentences that are literally inconsistent.

The problem becomes deeper, moreover, when we look to the importance inconsistency plays in expressivist accounts of validity, and hence solutions to the Frege-Geach problem. Again, expressivists cannot appeal to truth preservation or entailment to explain the validity of an argument. Instead, it has standardly been relations of consistency and inconsistency among the attitudes expressed that play the role. If normative sentence B follows validly from sentence A, a very natural explanation of this is that the attitude expressed by A rules out *alternatives* to the attitude expressed by B. But reasoning from ‘I don’t believe it is raining,’ to ‘It’s not raining’ cannot be a valid inference. Similarly, akratic agents are one of the paradigm cases of agents at odds with themselves. But reasoning from ‘I ought to exercise,’ to ‘I intend to exercise,’ or ‘I will exercise’ cannot be a case of valid inference. Not every case where we rationally commit ourselves to one attitude or another could, in

Schroeder's phrase, "license inference."

His solution to both problems is to restrict the type of psychological inconsistency to which his form of expressivism will appeal:

[Mark van Roojen's (1996)] diagnosis was that though Blackburn may have explained why there is something irrational about accepting the premises of an argument and denying its conclusion, irrationality is too easy to come by in order to suffice for an account of validity. ...

The kind of irrationality involved in accepting both 'Colorado is rectangular' and 'I don't believe that Colorado is rectangular' is often called *Moorean* inconsistency... Moorean inconsistency contrasts with the *genuine* inconsistency between 'P' and '~P'.

... [I]f expressivists are going to explain a version of the inconsistency property or the inference licensing property that suffices to distinguish valid argument from invalid arguments ... they need to appeal to the very kind of clash that obtains between beliefs with inconsistent contents.

(2010a: 122)

Schroeder names the form of inconsistency holding between beliefs with inconsistent contents *A-type inconsistency*.⁹ It is this restricted form of inconsistency Schroeder maintains that expressivists must appeal to: the inconsistency that holds

⁹ This restriction is challenged on various different grounds by (Sinclair 2011; Charlow 2014; Baker and Woods 2015; Silk 2015; and Alwood 2016).

between instances of the same type of attitudes with inconsistent contents. Note that not all attitudes are inconsistent when their contents are: it is plausible that one can desire p and desire $\sim p$ without incoherence, for example. On the other hand, to believe that p and believe $\sim p$ at the same time is incoherent, as is intending ϕ and $\sim\phi$ at the same time. We can call attitudes such as these, A-type. Thus we must stipulate that being for is, like belief, an A-type attitude.

The alternative to A-type inconsistency, B-type may be a form of psychological inconsistency, but it is not, for Schroeder, the type of inconsistency expressivists should appeal to if they are to account for literal contradiction, and thus solve the negation problem, or valid inference, and thus solve the Frege-Geach problem.

An interpretative point is worth briefly discussing here. (Baker and Woods 2015) read Schroeder as attempting to explain *logical* inconsistency and *logical* disagreement in terms of a special type of psychological conflict, namely A-type inconsistency. A similar reading is found in (Charlow 2014 and Silk 2015). This reading is supported by Schroeder's references to "logically related" and "logically unrelated" *attitudes* (see 2008b: 44-9, and 60). However, in recent discussion Schroeder has explicitly denied that reading (which raises the charge that his account of logic is overly psychologistic),¹⁰ and points to passages where he writes that logical

¹⁰ See the discussion at the *PEASoup* blog for Schroeder's denial that he has identified logical relations with psychological ones (<http://peasoup.typepad.com/peasoup/2015/01/ethics-discussions-at-pea-soup-baker-and-woods-how-expressivists-can-and-should-explain-inconsistency-with-precis-by-schroede.html>). The charge that Schroeder's account of logic is overly psychologistic is argued for in (Baker and Woods 2015).

relations between sentences, in addition to expressing A-type psychological relations, must also have a formal dimension (2008b: 68ff; 112).

Plausibly, then, Schroeder intends A-type inconsistency to explain broader forms of disagreement, inference, and argument than the merely logical. Starting first with inference and argument, A-type inconsistency is the property that explains why certain inferences are valid—but a distinction may still be made between *formally* valid inferences and *informally* valid arguments.¹¹ Schroeder need only identify the former as cases of logical inference. Logically valid arguments, on this reading, are those arguments that can be identified as valid solely in virtue of their logical form. Arguments are generically valid, moreover, just in case the attitude expressed by the conjunction of the premises is A-type inconsistent with the attitude expressed by the negation of the conclusion (2008b: 70ff.). Similarly, expressed A-type versus B-type inconsistency seems to be how Schroeder will avoid conflating semantic inconsistency with merely pragmatic inconsistency.¹²

4. The Objection

Remember, the real problem with Gibbard's treatment of negation is that he cannot

¹¹ As Schroeder puts it: "So let's take this as our definition of valid arguments: an argument is valid just in case anyone who accepts its premises is committed to accepting its conclusion, and logically valid just in case it is of a form all instances of which are valid" (2008b: 112).

¹² Of course, it is a necessary condition on an argument's logical validity that it be valid, and a necessary condition on logical inconsistency that the inconsistency be literal. So logical inconsistency will only hold between sentences expressing A-type inconsistent attitudes. But this on this reading, A-type inconsistency is necessary to but not sufficient for logical inconsistency.

distinguish internal and external negation of sentences deploying flavorless predicates. To fully evaluate the merits of Schroeder's proposal, we need to know how he would address these predicates. The obvious way for Schroeder to do so is to follow the same pattern as other normative predicates. We can say that 'X-ing makes sense' expresses:

FOR (ψ (x))

In which case 'Murder is wrong but not blaming for murder makes sense' expresses attitude (U):

(U) FOR (blaming (murder) & ψ (\sim blaming (murder)))

Why is this a problem? Well, why become an expressivist? One standard motivation is the widespread intuition that an agent who sincerely judges that something is the thing to do will be motivated to do it, barring weakness of will or other forms of defective agency. Someone who judges that some feeling is the one that makes sense suffers from a case of irrationality (of recalcitrant emotion) if they do not in fact feel it. Expressivism promises to explain this unusual connection between judgment and choice and feeling, by making the relevant predicates express motives or feelings—or attitudes with the role of regulating motives and feelings.¹³

¹³ This is the question of whether an expressivist should hold that normative sentences directly express motives and feelings, or higher-order attitudes towards motives and feelings.

Now, according to our proposed analysis, the agent who utters the above sentence is expressing a *commitment* (not a case of weakness or compulsion) to blame for murder, while simultaneously the same agent judges it makes sense not to blame for murder—and since ‘makes sense’ is normative, she must also be expressing a commitment *not* to blame for murder. So it seems like we should want to say that ‘Murder is wrong but not blaming for murder makes sense’ is semantically incompetent or she is irrational—she’s committing to doing something and simultaneously committing to not doing it. (U) must be, in some sense, an inconsistent state of mind.¹⁴ The question is, does Schroeder’s theory predict an inconsistent state of mind? If it doesn’t, we seem to have given up a standard reason for being an expressivist. If saying ‘Not blaming for murder makes sense’ (or is ‘reasonable’ or ‘appropriate’ or ‘One ought not to blame for murder’) doesn’t commit me, on pain of irrationality or semantic incompetence, to avoiding (in some manner) blaming for murder, then expressivism turns out to deny something we initially hoped it would explain, at least about flavorless predicates.

Just to be clear, the claim is not that the utterance ‘Murder is wrong, but one

See (Toppinen 2015) for discussion of this issue. Schroeder’s theory would be an example of a higher-order theory. Also see (Ridge 2015) for argument that the expressivist might not be as well positioned to explain the relevant datum as she thinks—and that in fact some form of hybrid expressivism scores better on this issue.

¹⁴ Note that the nature of the problem is conditional: *if* ‘wrong’ expresses a commitment to blame, *then* ‘Murder is wrong but not blaming for murder makes sense’ expresses a counter-normative commitment. But on Schroeder’s account, there must be some relation ϕ expressed by ‘wrong’, and so there must be a sentence ‘Murder is wrong, but it makes sense not to stand in ϕ toward murder’ that expresses an irrationality. Note also that this statement of the objection passes over conflicts between which attitudes are fitting or warranted, and which attitudes one ought to have all-things-considered (see footnote 3). But we can also construct sentences of the form ‘X is wrong, but murder does not warrant blame’, recreating the problem.

ought not to blame for murder' sounds infelicitous; I am happy to acknowledge it does not. The claim is simply that the agent must be expressing inconsistent psychological commitments, *on the assumption that* normative assertions get their semantic properties by expressing commitments. We can put the point this way: either the agent does not blame for murder, in which case she fails to have an attitude she is committed to having; or she blames, in which case her emotions don't conform with her own better judgment, her judgment about which reaction would make the most sense. So she is guilty of inconsistency in her psychological commitments.

But notice that with (U), at least, the conflict cannot be assimilated to A-type inconsistency. 'Murder is wrong' expresses:

(U1) FOR (blaming (murder))

Whereas 'Not blaming for murder makes sense' expresses:

(U2) FOR (ψ (\sim blaming (murder)))

(U1) and (U2) are not A-type inconsistent. So the defectiveness of (U) cannot be explained in terms of A-type inconsistency either.

On the other hand, the sentence 'Murder is wrong but one ought not to blame for murder' could be, within Schroeder's framework, analogous to a Moore-paradoxical sentence. Let's say 'Murder is wrong' and 'Not blaming for murder

makes sense' expresses attitudes B-type inconsistent with each other. Consequently, the agent who would utter the conjunction of the two is not guilty of a logical or narrowly semantic error (even if she is guilty of an irrationality in some broader sense). This is unproblematic, however, since the sentences in question do not logically contradict each other. Rather, she is expressing some broader form of incoherence in her attitudes.

This solution allows us to preserve the *explanandum* motivating expressivism: there is some sort of misunderstanding involved in stating that one ought to x , while also sincerely expressing a commitment not to x . But notice what we're really doing here: the being-for attitude is a theoretical posit. It was already stipulated that the attitude is conative and can stand in relations of A-type inconsistency. Now it seems we are stipulating further properties of it—namely which instances of it are B-type inconsistent.

Simply stipulating that (U1) and (U2) are B-type inconsistent would be an extraordinary cost to the theory, however. As noted earlier, normative judgments seem strangely connected with choice and motivation in a way that other judgments are not. This raises the question of why normative judgments and concepts are so different from the non-normative. Here is one *very unsatisfying* answer a cognitivist about moral judgments might give. Normative beliefs are different from other kinds of beliefs. The belief that X -ing makes sense is in rational conflict with the intention not to x .

This answer, however, offers no explanation of why normative beliefs or concepts are different; to claim that normative judgments play a different role in our

psychologies is to simply state what the problem is.¹⁵ Expressivism's selling point is supposed to be that it provides an explanation of that difference. Very simply, the attitude expressed by a normative sentence is a motivating attitude. So the agent who says 'Not *x*-ing makes sense' but chooses *x* is either insincere, or linguistically incompetent, or in psychological conflict with herself (she might have a plan to refrain from *x*-ing, expressed by her sentence, and another motive that leads her to fail to realize her plan).

The expressivist cannot claim to have achieved any of these explanations, however, if he simply stipulates that (U1) and (U2) are in rational conflict with each other. This is no different than postulating a brute rational connection between ought-beliefs and intentions. In other words, a stipulative solution preserves the *explanandum* at the cost of abandoning the promise of a simple *explanans*. We get half of the motivation for expressivism at the price of the other half.

I wish to make the nature of the objection clear. Given Schroeder's basic framework, he can explain why the predicate 'wrong' is attitude-guiding. 'Wrong' expresses being for blaming, and the being-for attitude is a conative, guiding attitude. That much is part of what we initially assume. The problem is that terms such as 'ought,' 'makes sense,' 'rational,' 'reasonable,' 'fitting,' and 'appropriate' are presumably action- and attitude-guiding as well. Now there is a sense in which Schroeder's framework can account for the action- and attitude-guiding role of the flavorless terms. But it is not a sense that is obviously adequate.

¹⁵ None of this should be taken to imply that realists do not explain the connection between ought-judgments and choice. For example, see (Smith 1994 and 1995; and Wedgwood 2007).

If I judge ‘Not blaming for murder makes sense,’ what this judgment should intuitively regulate is my attitude of blame (or my lack of such an attitude). If I continue to blame for murder, I am acting, by my lights, irrationally. But what the semantics actually predicts is that the judgment guides my second-order attitude (or other relation) of ψ -ing toward the absence of blame (Schwartz and Hom 2015: 839ff). It is this second-order relation that the judgment guides; and a failure to have this relation is a breakdown in rational incoherence, or so the theory would predict. By analogy, the judgment that ‘Not blaming for murder is good’ would express the attitude of, for example, being for *desiring* not to blame for murder.¹⁶ And failure to have that desire will presumably be a failure of rational coherence—given the commitment (of being for) to have the relevant desire. But it is an open question whether there is any breakdown in rational coherence in failing to have an attitude one desires to have—and thus an open question whether there would be anything irrational about continuing to blame for murder despite having sincerely judged that it would be good not to.¹⁷

¹⁶ The assumption here is that if *better than* expresses preferring (2008b: 58), then *good* should express desiring.

¹⁷ In their (2015) Schwartz and Hom raise this same problem (or a very similar one) with Schroeder’s semantics. As they put it: “It is important to note that [for Schroeder] ordinary judgments about some activity always correspond to a *being for* of some relation to that activity. It is thus impossible to simply *be for* giving to charity...” (839). I have a few points of disagreement with some of their development of this objection, however: two minor and one more significant.

First, I do not think the text supports the claim that for Schroeder *being for* giving to charity is impossible. (Schroeder may hold this, but his published work simply leaves the issue open.) Rather, what Schroeder is committed to is the claim that *being for* giving to charity is not expressible through a declarative sentence that predicates some property or relation of a subject. It could be that such attitudes exist, however; they are simply inexpressible or else expressed through non-declarative sentences with simpler logical structure, such as exclamations or imperatives (Köhler 2012 shares this reading of

This is as it should be: one could judge that it is good not to blame for murder, because giving up on blame leads to greater mental health, while still holding that blaming murderers *makes sense* or *is fitting* or *is the thing to do*. The problem is, we can also judge that not blaming murderers makes sense (or is fitting, etc), and *that judgment* does rule out the rational coherence of continuing to blame for murder. So our account of ψ must not leave it an open question whether continuing to blame for murder, after one has judged that it is not doing so that makes sense, is coherent. Flavorless predicates favor or disfavor their objects, we might say, *directly*. Whereas flavored predicates plausibly favor taking certain attitudes (or standing in other relations) to their objects.

Admittedly, one might object at this point that Schroeder's semantics is based on the presupposition that no normative predicates favor or disfavor their

Schroeder). Schwartz and Hom are correct however that no normative declaration can express simply *being for* giving to charity.

Second, Schwartz and Hom argue that Schroeder may be forced to give up one of the standard explanatory advantages of expressivism, "a compelling story on how accepting a moral claim can be motivating" (840). Notice, however, that they assume here that the expressivist wants to tell story on which moral judgments motivate performing or refraining from the morally valenced action. But among Gibbard's grounds for analyzing 'X is wrong' as expressing a plan to feel anger to those who X was a desire to accommodate the possibility of agent's who hold that it may in some cases be irrational to act morally. (For more discussion of this possibility, see footnote 19). It is arguably an attractive feature of Schroeder's theory, then, that it does not make all normative predicates favor their objects directly. The problem, I think, is that we should want at least *some* predicates to favor their objects directly (as on Gibbard's semantics).

Third, and most importantly, Schwartz and Hom seem to assume that the primary way in which a normative attitude might favor an object directly is simply for the attitude to have that object as its content, rather than a relation to that object as content (i.e., an instance of *being for* would directly favor giving to charity if it were an instance of *being for* giving to charity). I think this overlooks a possibility. I believe there is a way for an instance of *being for* ψ -ing giving to charity to directly favor giving to charity, in virtue of favoring standing in relation ψ to giving to charity: namely, if *being for* ψ -ing giving to charity closes off further question about whether to give to charity; or, in other words, if it is irrational to *be for* ψ -ing giving to charity, but still fail to give to charity. I develop this possibility in the following discussion.

objects directly. Normative predicates contribute some relation which one favors standing in (i.e., one is *for* standing in that relation) to the object of evaluation (2008b: 58).¹⁸ To insist that flavorless predicate favor or disfavor their objects directly is to insist on something the semantic picture denied from the beginning. This statement of the problem, then, risks begging the question, or so the objection would go.

The first thing to say in response to the charge of question-begging, is to point out that the case just made for the status of flavorless predicates as directly favoring their objects was not that Schroeder seems committed to this, nor some assumption that Schroeder's semantics should be more like Gibbard's or anyone else's. Rather, it is simply the assumption that *some* normative predicates settle the matter about whether to perform their objects (when their objects are actions) or settle whether to realize their objects (when their objects are attitudes). Let's assume, very plausibly, that murder is wrong. There is still the question (it's at least seemed to many) of whether to commit murder anyway. This is the question of whether murder might sometimes be reasonable, of whether the reasonable agent will always be a moral one, or whether non-moral considerations, considerations of self-interest, aesthetic value, authenticity, and so on, sometimes outweigh moral considerations. On the other hand, it is generally assumed that there is some predicate that, when applied to the object *murder*, no longer leaves open the question of whether to murder. The term in question may be 'reasonable', as we've assumed, or 'the thing to do', or 'makes sense', or something else. But this term, if it is to serve to close further

¹⁸ Thanks to a referee for raising this point.

questioning about whether to murder, must directly favor or disfavor its object, in a way that ‘wrong’ plausibly does not.¹⁹

Schroeder could perhaps deny that there is any predicate that directly favors its object. However, as we have seen, this would be to deny that there is any normative predicate that closes off further question about whether to perform some action. It would also deny that there is a predicate that closes off whether to have some attitude, when the attitude figures in the judgment as the object of endorsement or condemnation. (Presumably ‘Murder is wrong’ closes the question of whether to blame for murder; but without a predicate directly favoring its object there could not be a sentence of the form ‘Blaming for murder is *F*’ which similarly

¹⁹ A referee asks whether ‘Murder is wrong, but it’s wrong to blame for murder’ expresses inconsistency—perhaps of a pragmatic rather than semantic form. After all, the attitude expressed by this sentence would be:

FOR (blaming (murder) & blaming (blaming (murder)))

The speaker is expressing an attitude which she is also expressing a commitment to blame.

The answer, I think, is that it is not inconsistent by itself, precisely because it is an open question whether moral judgments are overriding, or whether a commitment to blaming for X rationally commits one not to X. If moral judgments are overriding, such that *Murdering is wrong* entails that *Not murdering makes sense*, then ‘Murder is wrong but blaming for murder is wrong’ is inconsistent. But notice that in this case the speaker is also rationally committed to having attitude (U). This is the source of the inconsistency.

On the other hand, if moral demands are sometimes superseded by other kinds of considerations, there is no inconsistency. Consider the person who says ‘Murder is wrong, but it’s wrong to blame for murder.’ She is committing to blaming for murder. She is also committing to blaming those who blame for murder (including herself). So she is committed to doing something she is committed to feeling guilty about (assuming the first-person expression of blame is guilt). This may seem weird, and possibly irrational, but remember that if moral requirements are not always overriding there will presumably be cases like this. Take Williams’ (1981) version of Gaugin: one could conclude that Gaugin is doing something morally wrong in abandoning his family to pursue art, and so it is reasonable for him to feel bad about his choice. But values of artistic creation, authenticity, and so on win out in this case, and so it is reasonable for him to make that choice even while he feels bad about it. If cases like this are possible, there should be no necessary irrationality in committing to having *attitudes* that one will appropriately feel bad about having either.

settled the issue.)

This position is not absurd. It is similar to certain views that deny “the unity of reason,” (Copp 1997; Tiffany 2007; Baker, *ms.*) holding that a variety of valid normative stances and perspectives can be taken to some action or attitude, but none of these evaluations identify the action in question as the thing to do (or at least they fail to identify it as the thing to do *full stop*; they might mark some response as the thing to do *morally*, or the thing to do *self-interestedly*). The arguments for and against this position are too involved to summarize here. I will simply assume that it is not Schroeder’s position, because whatever its merits, it is a very radical one. Presumably, if Schroeder intended for his semantics to have consequences such as these, they would be explicitly addressed. Of course, I may be wrong in this, in which case much of the subsequent argument is mistaken. But the argument thus far will at least establish a surprising consequence of Schroeder’s theory: it implies that an unorthodox metaethical theory must be true on semantic grounds.

If Schroeder were to abandon flavorless predicates, it would also change the nature of his objection to Gibbard considerably. Remember that Gibbard can provide sufficient structure for distinguishing forms of negation for flavored predicates. Gibbard’s mistake, it would turn out, is not that he lacks the resources for accounting for flavorless predicates; it’s that he tried to account for them.

Besides these points, we can answer the charge of begging the question by offering a neutral test, compatible with Schroeder’s semantics, which, if met, would amount to identifying a meaning for the predicate ‘makes sense’ that allowed the term to function as a direct endorsement of its object, or something close enough to

a direct endorsement to satisfy the sense that this term should close off further question about what to do or to feel. The person who utters ‘Murder is wrong’ is committed to blaming for murder. If the predicate ‘makes sense’ does indeed close off further question about what to do, the person who sincerely utters ‘Not blaming for murder makes sense’ is committed to not blaming for murder. So we need to give an account of ψ that predicts rational conflict between (U1) and (U2). This would amount to an account of ψ on which (U2) offered something close enough to direct guidance to satisfy the sense that some normative predicate must directly favor its object.

But there is a caveat. If this prediction is only bought by stipulation, then we have accounted for the action- and attitude-guiding role proper to flavorless predicates by fiat—and anyone can do as much. Yet it is impossible, I will argue, to find a ψ that allows for an explanation without incurring significant costs elsewhere.

It is worth remembering at this point that Gibbard can avoid this problem because he *defines* flavored predicates in terms of flavorless predicates. Thus, ‘Not blaming for murder makes sense’ and ‘Murder is not wrong’ both guide feelings of blame in the same direct way for Gibbard, because the sentences are equivalent in meaning. This, of course, leads to the problem of negation for Gibbard. Schroeder can solve this problem with the semantics, but only at the cost of making mysterious the relation between wrong-judgments and judgments about whether blaming makes sense.

4.1. The Problems with Identifying ψ with Being For

Before continuing, it's worth noting why this problem will be very hard to solve. *Being for* is, to repeat, a theoretical posit. Virtually everything we know about the attitude is stipulated; we can have only the most minimal intuitions about how it functions and what rational relations it stands in. That massively restricts the theoretical resources the attitude offers us for providing any principled explanation. But we don't want a stipulative solution for why (U1) and (U2) are in rational conflict.

Is there anything we can assume about being for that isn't stipulative? As noted earlier, it seems like an implicit commitment about the nature of the being-for attitude is that failing to have the target attitudes that one is for having is a failing of rational coherence.²⁰ Presumably it was always part of the story that an agent who is for blaming for murder but does not in fact blame for murder is guilty of irrationality (all else being equal).

As was also noted, some conative attitudes may not be like this; in fact, I will argue later that some cannot be. But it seems an assumption of Schroeder's semantics that being for is. This suggests one obvious way in which (U1) and (U2) are inconsistent: ψ might just be the being-for attitude. (U2) would thus be a third-order commitment of being for being for not blaming for murder. This commitment would be satisfied if the agent is for not blaming for murder. But this target second-order commitment is straightforwardly A-type inconsistent with (U1)—which is also a second-order commitment. On the other hand, if she does not satisfy (U2), she is

²⁰ But see (Toppinen 2015) for criticism of higher-order attitude versions of expressivism.

guilty of rational failing because of the failure of satisfaction. So, if an agent has both attitudes (U1) or (U2) she is necessarily irrational: she is irrational if she is for not blaming for murder, and irrational if she is.

This solution is straightforward. Unfortunately, it brings unwelcome theoretical commitments. Here is a reasonable sentence: ‘It makes sense to jog every morning’. If ψ is identical with being for, the sentence must express the following:

FOR (for (jogging every morning))

But then the being for attitude can take *jogging every morning* as its content. In other words, the being for attitude can take unstructured objects for its content; not all of its contents must have the structure $\phi(x)$. This raises the question of which sentence is expressed by the following:

(J) FOR (jogging every morning)

The problem of which sentence expresses an attitude such as (J) has already been raised in (Skorupski 2012), and answered in (Schroeder 2012; and Köhler 2012). So it’s worth looking at the replies given. First, nothing in Schroeder’s earlier work commits him to holding that an attitude such as (J) is possible. Schroeder could hold that necessarily any case of being for must take a content with the structure $\phi(x)$. This is a fine response to Skorupski. But if we are identifying ψ with being for, it is

unavailable.

Another possible reply is that attitudes such as (J) are not linguistically expressible. Only cases of the being for attitude with properly structured content can be expressed in sentences (Köhler 2012). This response is, however, highly problematic. (J) is A-type inconsistent with the following:

(NJ) FOR (\sim jogging every morning)

So a person with the attitude (J) would be in logical disagreement with the person with the attitude (NJ)—disagreement in the exact same way that two people who accept literally inconsistent attitudes are in disagreement. But they would not be able to put this disagreement into words—or at least, not into words other than the ones just used. This is strange. Likewise, one could derive other attitudes from (J) or (NJ) through a process of logical inference. An agent could have the following attitude:

(JS) FOR (jogging every morning \vee swimming every morning)

If a person has attitudes (JS) and (NJ), presumably being for swimming every morning would follow, and it would follow in just the same way as believing that p follows from believing that p or q and believing that not q . In other words, it seems to follow via valid inference. But again, this inference, despite being valid, could not be represented as an argument with premises and a conclusion.

The problem gets even worse. Our moral commitments—which are

assertible in natural language—could easily follow from commitments such as (J).

For example, suppose I have the following attitude:

(JW) FOR (jogging every morning \rightarrow \sim blame (jogging every morning))

And so one could validly infer the correctness of ‘jogging every morning is not wrong’ from (JW) and (J), but this inference would not be expressible. This means that an agent could have something like grounds or reasons for her conviction that jogging is not wrong, but she could not state any of these to defend her conviction.

Besides defending our own convictions, we use normative argument to get people to see what follows from their commitments. Cases of being for with unstructured objects also allows the possibility that some moral conviction follows, through a process of valid inference, from someone’s other existing commitments—and yet it would be impossible to use natural language arguments to get the person to recognize this.

It is a mystery, then, why natural languages have not lexicalized the being-for attitude, given the obvious utility of doing so. One could be in a debate about the wrongness of murder (or any other normative matter), and be unable to express in natural language a rational defense of one’s opinion, even if one had commitments that made the opinion eminently defensible. One’s disputant could have failed to draw the conclusion that follows from his own commitments, but there would be no way in English to make this apparent to him. Being able to do these things would be very useful. Why has our language not developed to let us do so?

To be clear, Schroeder is not committed to allowing the possibility of attitudes such as (J), and so his semantics does not run into these problems. This is an objection to a particular way of developing Schroeder's semantics in answer to a problem: the proposed extension breeds mysteries.

It is true of course that there are many attitudes we cannot express in language. And there is a general mystery for expressivists of why we can express some in natural language and not others; and it is unclear if this is stranger than any of the mysteries surrounding natural language generally, whatever one's larger semantic picture. The objection here is not that general problem, however. It is the specific problem that, once we grant that instances of being for with structured contents are expressible, it becomes mysterious why instances with unstructured commitments are not, given the utility such sentences would have in coordinating normative judgments.

(Köhler 2012) suggests another way of dealing with attitudes such as (J): (J) and (NJ) are expressible, but not through declarative sentences. Köhler does not elaborate on which non-declarative sentences might be doing the work, but in relation to the problem of inexpressible logical disagreement, the natural answer would be imperatives. Our attitudes could be expressed through 'Jog every day!' and 'Don't jog every day!' This potentially brings another advantage to Schroeder's semantics: (Charlow 2014) has pointed out that imperatives, such as those just given, seem to have enough logical structure that they can logically contradict one another, and argues that Schroeder's assumptions about the nature of logical inconsistency seem to rule this out. So embracing an imperatival interpretation of attitudes such as

(J) could potentially solve two problems at once.

The problem is that imperatives can also take relations to objects as their content; for example, ‘Don’t blame for murder!’ or ‘Believe the truth!’ We would now need an interpretation of these sentences. We could say that the first sentence is semantically equivalent to ‘Murder is not wrong’. But then we must explain why the latter sentence can figure into a range of embeddings that the imperative cannot, why the truth predicate intelligibly applies to the latter but not the former, and so on.

Also notice that on the above proposal, the following attitude can be expressed:

FOR (blaming (murder) \rightarrow \sim murder)

The sentence expressing it would be, presumably ‘If murder is wrong, then don’t murder!’

But this attitude is inexpressible:

FOR (\sim murder \rightarrow blaming (murder))

Perhaps this seems right. After all, the sentence “If don’t murder(!) then murder is wrong” is nonsense. But notice this means again that certain logically valid inferences are inexpressible as arguments. For example, an agent with the attitude of *being for* (\sim murder \rightarrow blaming (murder)) and the attitude of *being for* (\sim murder), is committed to murder being wrong, on the assumption that valid inferential relations are fully

explained in terms of A-type inconsistency. So one could have commitments from which the wrongness of murder is a valid consequence, but be unable to offer the grounds for one's conviction. Again, this is bizarre.

One final option (which again draws from points found in Köhler 2012)²¹ would acknowledge disagreement between (J) and (NJ), but insist that they are merely inexpressible *in natural language*. But we could invent some artificial language that does express it.²² Once again, though, this raises the question of why these attitudes are currently inexpressible, given how useful expressing them would be.

Perhaps some additional considerations can be given to develop one of these solutions into a compelling answer—explanations of why, despite the obvious utility of lexicalizing being for, natural languages do not, or arguments that this is not as surprising as it seems. But in the absence of one, identifying ψ with being for looks unattractive.

4.2. Identifying ψ with Any Other Attitude

What about identifying ψ with some other attitude? This is not particularly hopeful either. Let's call ψ , whatever it is, *endorsement*. Now endorsement must either be an A-type attitude, or not.

²¹ This is not to say that Köhler would endorse any of this. I mean only to credit him for the good; I take all responsibility for the bad.

²² Schroeder, for example, introduces the FOWR predicate to express the attitude of being for (2008b: 60-4). As we will see, the problem is not with any of the logical or semantic properties of this artificial term, but the question of why we had to introduce it in the first place. Thanks to a referee of an earlier draft for pointing to this issue.

If it is not an A-type attitude, it is unclear how we could explain the (B-type or Moorean) inconsistency between (U1) and (U2). Recall that earlier we said that it is plausible that it is a violation of rational coherence to fail to satisfy certain conative commitments because one's own mind is in some sense uncooperative: for example, it seems like it must be part of the story about the being-for attitude that being for blaming for murder, but then failing to blame for murder, is a violation of coherence. But this kind of coherence relation could *only* hold of A-type commitments. After all, assume that it is psychologically inconsistent to allow the endorsement of x go unsatisfied. Then an agent who endorses x -ing and endorses $\sim x$ -ing must be inconsistent. Either he leaves the first unsatisfied (which is inconsistent), or he leaves the second unsatisfied (which is inconsistent). But then endorsement must be an A-type attitude.

Put another way, if we assume that there is nothing inconsistent about endorsing x and endorsing $\sim x$, it is completely mysterious how there could be something inconsistent about endorsing x , but then bringing about $\sim x$.

So if endorsement is not A-type, we cannot appeal to failures to satisfy one's commitments to explain the B-type inconsistency that obtains between (U1) and (U2). After all, one could satisfy (U1) by blaming for murder. One could satisfy (U2) by endorsing not blaming for murder. And there is no inconsistency in failing to satisfy endorsements. All the attitudes that demand satisfaction are satisfied; those unsatisfied do not demand satisfaction.

On the other hand, if endorsement is A-type, we fall back into the problem of inexpressible disagreement. Endorsement must be either primitively A-type

inconsistent, or, like beliefs and possibly intentions, it reduces to a case of A-type inconsistency between attitudes of being for. If the former, then agent who endorses murdering and the agent who endorses not murdering are in logical disagreement with each other, but they cannot express the disagreement, since sentences express cases of the being for attitude. Or, if instances of endorsement could be expressed by sentences, we face a new problem. Let's say 'P' expresses being for x and 'Q' expresses endorsement of y . Notice that there is no way now of preserving the standard logical relations between 'P', 'Q', and ' $\sim P$ or $\sim Q$ '. Those sentences are logically inconsistent, but the attitudes expressed cannot be A-type inconsistent, since 'P' and 'Q' express different attitude-types. We could deny that mixed-expressions such as ' $\sim P$ or $\sim Q$ ' are possible, but then we are committed to finding some portion of natural language that is effectively partitioned from the rest of language with respect to compositionality. Or we could insist that the above triple is not really logically inconsistent: but then we are committed to massive ambiguity in our logical vocabulary.

One might try to get around this problem by suggesting that endorsement is expressed by imperatives. But imperatives can figure in conjunctions and conditionals with non-imperative sentences, albeit more restrictedly. Sentences such as 'Murder is wrong, but don't blame for murder!' or 'If murder is wrong, then don't murder!' are meaningful. Treating endorsement as a primitively A-type attitude looks unattractive, then.

Suppose instead that endorsement reduces to a case of being for. Just as believing that p is identical with being for proceeding as if p , endorsing x is identical

with being for ξ -ing x , for some ξ . This introduces a familiar and (as Schroeder makes clear in *Being For*) recurring problem with expressivist semantics: we now have the wrong number of negations, in this case *too many*. Consider the following sentences:

- (O1) One ought to jog.
- (O2) One ought not to jog.
- (O3) It's not the case that one ought to jog.

These will express the following attitudes:

- (A1) FOR (for (ξ (jogging)))
- (A2) FOR (for (ξ (\sim jogging)))
- (A3) FOR (\sim for (ξ (jogging)))

But there is one additional possible attitude, apparently not expressible by any ought-sentence (or sentence using a predicate of 'reasonable'):

- (A4) FOR (for ($\sim\xi$ (jogging)))

But (A4) is A-type inconsistent with

- FOR (\sim for ($\sim\xi$ (jogging)))

Once again the problem of inexpressible disagreement—and by implication, inference and justification—looms.

Schroeder encounters the problem of too many negations in his account of belief (2008b: 95ff), and so it is natural to suspect that he could solve this problem in the same way. The paper will present a very simplified version of his solution, *bifurcated attitude semantics*, to show that its extension to this case is problematic. (The simplification is necessary due to constraints of space, but it will not impact the argument here. As we will see, the reason Schroeder’s solution cannot be extended is itself very simple.)

The belief that p , remember, expresses being for proceeding as if p . So for any sample sentence ‘P’, we have a single corresponding negation of that sentence ‘ \sim P’. But the content of the attitude of being for proceeding as if p (i.e., believing that p) has two places where it can be negated:

(N1) FOR (proceeding as if (\sim p))

(N2) FOR (\sim proceeding as if (p))

Schroeder solves this problem by introducing bifurcated-attitude semantics. These semantics explicitly depend on “a simple logic of proceeding as if,” according to which “proceeding as if p entails not proceeding as if $\neg p$ ” (2008b: 98). This logic is justified, for Schroeder, by the assumption that “it is not possible to both proceed as if p and simultaneously proceed as if $\neg p$ ” (97). *Simplifying a great deal*, this solves the

problem because proceeding as if p entails not proceeding as if $\sim p$, and so being for the first state commits an agent to being for the second; the sentence expressing (N1) thus also expresses (N2), since (N2) is a commitment following from (N1).²³

So, if we posit that ξ -ing $\sim x$ entails not ξ -ing x , it will turn out that the sentence expressing (A2) will also express (A4), since the latter commitment simply follows from the former (i.e., what one commits to with attitude A2 has identical conditions of satisfaction with what one commits to with A4). The problem here is that by postulating such a relation to explain the meaning of ‘makes sense,’ we end up pushing the bump in the explanatory rug once again. After all, if there are commitments of the form

FOR (for (ξ (jogging)))

then these commitments are only satisfied if the agent in question is in fact FOR (ξ -ing (jogging)). But then there must be some predicate expressing ξ , otherwise again we have inexpressible disagreement and inexpressible inference. Unfortunately ξ comes with very specific theoretical commitments. *ξ -ing not jogging* entails *not ξ -ing jogging*. But this means we need a predicate for which $F(\sim x)$ entails $\sim F(x)$, and which is also a plausible candidate to explain what it is we are expressing being for when we say that a choice or attitude ‘makes sense.’ But which attitude is this?

²³ This summary skips over major and minor attitudes, the attitude of disbelief, and other complexities. Interested readers should see chapters 7 and 8 of *Being For*.

4.3. Identifying ψ with Some Non-attitudinal Relation

In some cases of being for with structured content the structure is provided not by a target attitude but by some other type of relation: *proceeding as if* in the case of descriptive sentences, for example. So perhaps we can identify ψ with some relation that is not an attitude.²⁴

But which relation? I will canvass the two most obvious. First, *ψ -ing* could be identical to *recommending*. This does seem to capture something of the quality of flavorless endorsement. The problem is, there isn't any obvious conflict between (U1) and (U2) on this interpretation of ψ . A policy of recommending against blame while committing to blaming may be hypocritical, but there is no way in which my conflicts put me at odds with myself, unless I have further commitments not to advise others to do what I would be unwilling to do myself. Our speaker could be a self-conscious hypocrite who is untroubled and unconflicted about her hypocrisy. In other words, *recommending* as an interpretation of ψ doesn't seem to produce the requisite psychological inconsistency.

Perhaps we could understand ψ as expressing *recommending to oneself*. But the idea of recommending to oneself is metaphorical, and the most obvious way of understanding the metaphor is in terms of a higher-order endorsement—that is, treating ψ as an attitude.

Another relation might be *instantiation*. If I judge that blaming for murder makes sense, perhaps what I am for is instantiating the blame-relation toward

²⁴ Thanks to a referee of an earlier version for pointing to this possibility.

murder; and if I judge that jogging makes sense, I am for instantiating the act-type of jogging. It is unclear if this conjecture is available to Schroeder, given that he uses the relation of instantiation to explain the meaning of the predicate ‘true’ (2008b: 158ff.). In any case, this suggestion will not work. If *recommending* was too indirect a relation to serve as ψ , *instantiating* is so direct it eliminates the distinction between internal and external negation. The only way to *not* instantiate blame for murder is to not blame for murder. But then we lose the original distinction we were after. ‘Blaming for murder does not make sense’ is now equivalent in meaning to ‘Not blaming for murder makes sense.’

The failure of these two cases suggests a general dilemma for any attempt to identify ψ with some non-attitudinal relation.²⁵ Either not ψ -ing blaming for murder entails not blaming for murder, or it does not. If the entailment holds, the original negation problem facing Gibbard reappears: we cannot assign distinct meanings to cases of external negation and internal negation in sentences using the flavorless predicate. On the other hand, if the entailment doesn’t hold, we still face the challenge of explaining why there would be a rational conflict between being for ψ -ing not blaming for murder and being for blaming for murder.

5. Conclusion

Note that Gibbard can easily explain the defect in ‘Murder is wrong but not blaming for murder makes sense’. He doesn’t even need to appeal to expressivism. Given the

²⁵ Thanks to an anonymous referee for pointing this out to me.

analysis of ‘wrong’ in terms of ‘makes sense’, the statement is a straightforward analytic falsehood, just as ‘Pat is a bachelor but Pat is a woman’ would be. Thus, the fact that the agent who sincerely utters this sentence must have incompatible commitments is unproblematic. Of course, the preceding objections to Gibbard’s account remain.

It should be acknowledged that Schroeder’s theory still represents a substantial advance in expressivist semantics. The problems raised here arise, I would argue, from the focus of his books on expressivism as a *semantic theory*. It is that, of course, but it is also a theory with consequences for metaphysics and moral psychology. Expressivism is committed to a much tighter connection between meaning and reasoning than would be true on more realist semantics (and note that “reasoning” here should be understood broadly enough to include practical reason that results in actions or feelings). Obviously any theory must grant that there is some. But in the case of Schroeder’s expressivism, the relations of entailment between claims hold in virtue of the inferential or rationalizing relations of the attitudes expressed. Sentence ‘B’ is a valid consequence of ‘A’ because the attitude expressed by ‘ \sim B’ is A-type inconsistent with the attitude expressed by ‘A’. It is thus a special class of the rationalizing relations between attitudes that explain semantic relations such as entailment or inconsistency holding between the sentences expressing them.

There are very natural pressures pushing the expressivist to accept as much. But then there is the problem that, in developing the semantics, we may end up postulating more (or fewer) of the relevant rationalizing relations than we want. For

example, if the being for attitude can take unstructured objects, the concern that there are valid inferences that cannot be expressed as valid arguments looms. In other words, if the *exact same type of reasoning* is involved in the management and updating of the unstructured instances of being for as in the structured, why are only the latter expressible? Why is some of our reasoning, of the very same type, silent—and consequently inaccessible to public debate or critique?

Similarly, there are many normative predicates, and we need some story about how they relate to one another, not simply semantically—but also how the judgments involved potentially conflict or potentially add up to all-things-considered judgments. ‘Wrong’ and ‘makes sense to blame for’ both seem to express a commitment to blaming. Is it the same kind of commitment? If they are different commitments, how do these interact? The plausibility of the semantic picture cannot be assessed independently of what it tells us about these psychological questions.*

Works Cited

- Alwood, Andrew (2016) “Non-Descriptive Negation for Normative Sentences,” *Philosophical Quarterly* 66 (262): 1-25.
- Baker, Derek (*manuscript*). Skepticism about Ought *Simpliciter*.
- Baker, Derek, and Jack Woods (2015) “How Expressivists Can and Should Explain Inconsistency,” *Ethics* 125(2): 391-424.
- Blackburn, Simon (1998) “Attitudes and Contents,” in his *Essays in Quasi-Realism*, Oxford: Oxford University Press, pp. 182-97.

* Thanks to Jack Woods and anonymous referees for extensive comments and criticism. Research appearing in this paper was partially funded by a grant from the Research Grants Council of Hong Kong, China (LU342612).

- Björnsson, Gunnar, Caj Strandberg, Ragnar Francén Olinder, John Eriksson, and Frederick Björklund (eds.) (2015) *Motivational Internalism*, Oxford: Oxford University Press.
- Charlow, Nate (2014) "The Problem with the Frege-Geach Problem," *Philosophical Studies* 167(3): 635-65.
- Copp, David (1997) "The Ring of Gyges: Overridingness and the Unity of Reason," *Social Philosophy and Policy* 14(1): 86-101.
- D'Arms, Justin, and Daniel Jacobson (2000) "Sentiment and Value," *Ethics* 110(4): 722-48.
- Dreier, James (2006) "Negation for Expressivists: A Collection of Problems with a Suggestion for a Solution" in R. Shafer-Landau (ed.) *Oxford Studies in Metaethics, Volume 1*, Oxford: Clarendon Press.
- Gibbard, Allan (1990) *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Harvard University Press.
- _____ (2003) *Thinking How to Live*, Harvard University Press.
- _____ (2012) *Normativity and Meaning*, Oxford University Press.
- Hieronymi, Pamela (2005) "The Wrong Kind of Reason," *The Journal of Philosophy* 102(9): 437-57.
- Horgan, Terry, and Mark Timmons (2006) "Cognitivist Expressivism," in T. Horgan and M. Timmons (eds.) *Metaethics after Moore*, Oxford University Press.
- Köhler, Sebastian (2012) "The Frege-Geach Objection to Expressivism: Structurally Answered," *Journal of Ethics and Social Philosophy*.
- McGeer, Victoria (1996) "Is 'Self-Knowledge' an Empirical Problem? Renegotiating the Space of Philosophical Explanation," *The Journal of Philosophy* 93(10): 483-515.
- Moran, Richard (2001) *Authority and Estrangement: An Essay on Self-Knowledge*, Cambridge, MA: Harvard University Press.
- Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen (2004) "The Strike of the Demon: On Fitting Pro-Attitudes and Value," *Ethics* 114(3): 391-423.
- Ridge, Michael (2014) *Impassioned Belief*, Oxford University Press.
- _____ (2015) "Internalism: *Cui Bono?*" in (Björnsson et al. 2015: 135-149).
- Schroeder, Mark (2008a) "How Expressivists Can and Should Solve their Problem with Negation," *Nous* 42(4): 573-99.
- _____ (2008b) *Being For*, Oxford University Press.

- _____ (2010a) *Non-Cognitivism in Ethics*, Routledge.
- _____ (2010b) "Value and the Right Kind of Reason," in *Oxford Studies in Metaethics, Volume 5*: 25-55; ed. R. Shafer-Landau, Oxford: Clarendon Press.
- _____ (2012) "Skorupski on *Being For*," *Analysis* 72(4): 735-9.
- Schueller, G.F. (1988) "Modus Ponens and Moral Realism," *Ethics* 98: 492-500.
- Schwartz, Jeremy, and Christopher Hom (2015) "Why the Negation Problem Is Not a Problem for Expressivism," *Nous* 49(4): 824-45.
- Sharadin, Nathaniel (2013) "Schroeder on the Right Kind of Reason," *Journal of Ethics and Social Philosophy* 7(3).
- Shiller, Derek (2016) "A Primitive Solution to the Negation Problem," *Ethical Theory and Moral Practice* 19 (3): 725-40.
- Shoemaker, Sydney (1996) *The First-Person Perspective and Other Essays*, Cambridge: Cambridge University Press.
- Silk, Alex (2015) "How To Be an Ethical Expressivist," *Philosophy and Phenomenological Research* 91(1): 47-81.
- Sinclair, Neil (2011) "Moral Expressivism and Sentential Negation," *Philosophical Studies* 152(3): 385-411.
- Skorupski, J. (2012) "The Frege-Geach Problem for Expressivism: Still Unanswered," *Analysis*, 72(1): 9-18.
- Smith, Michael (1994) *The Moral Problem*, Blackwell.
- _____ (1995) "Internal Reasons," *Philosophy and Phenomenological Research* 55(1): 109-31.
- Tiffany, Evan (2007) "Deflationary Normative Pluralism," *Canadian Journal of Philosophy* 37(5): 231-62.
- Toppinen, Teemu (2015) "Pure Expressivism and Motivational Internalism," in (Björnsson et al. 2015: 150-166).
- Unwin, Nicolas (2001) "Norms and Negation: A Problem for Gibbard's Logic," *The Philosophical Quarterly* 49: 337-52.
- Van Roojen, Mark (1996) "Expressivism and Irrationality," *The Philosophical Review* 105(3): 311-35.
- Wedgwood, Ralph (2007) *The Nature of Normativity*, Oxford: Clarendon Press.

Williams, Bernard (1981) "Moral Luck," in his *Moral Luck*, Cambridge University Press, pp. 20-39.

Woods, Jack (2014) "Expressivism and Moore's Paradox," *Philosophers' Imprint* 14(5): 1-12.

Wright, Crispin (1988) "Realism, Antirealism, Irrealism, Quasi-Realism," *Midwest Studies in Philosophy* 12: 25-49.

Zangwill, Nick (1992) "Moral Modus Ponens," *Ratio (New Series)* 2: 177-93.