

# Comments on Schroeder, *Being For*

Seth Yalcin\*

Feb 7, 2012

## 1 Possibility versus empirical plausibility

Metaethical expressivism is, *inter alia*, supposed to be a thesis about the meaning of some fragment of language. About this sort of thesis we can separate two questions:

1. LOGICAL POSSIBILITY. Is the thesis even *in principle* compatible with a compositional semantics for some elementary possible language (say, the language of propositional logic)?
2. EMPIRICAL PLAUSIBILITY. Can the thesis be well-motivated for a fragment of some actual natural language?

Given standard assumptions, in particular the assumption that natural languages are compositional, an affirmative answer to the second question entails an affirmative answer to the first question. I take it that expressivists are ultimately interested in defending an affirmative answer to the second question, and I take it a lot of the preoccupation in the literature has been with the first question *en route* to the second.

Concerning the first question, it is surely a banality that the answer to this question must be affirmative if we are to take any given expressivist thesis at all seriously. Bracketing some kind of general skepticism about the explanatory power of compositional semantics, the thought that a proposal partly about the meaning of some natural language needs at least to be compatible in principle with a compositional semantics is about as elementary and uncontroversial in contemporary theorizing about meaning as it gets. Commensurate with this, the expressivist shouldn't get too excited if she can show how to answer the first question affirmatively. For all that would mean is that she has met a very minimal prerequisite for being taken seriously. It would still remain to show that the kind of theory her view requires bears even a foggy resemblance to what English, or whatever, calls for.

---

\*yalcin@berkeley.edu.

## 2 Possibility

So much is familiar. Temporarily setting the issue of empirical plausibility to one side, let us consider the situation with the first question. Do we have on hand any metaethical expressivist package which can be seen to cohere with a compositional semantics for some possible language—say, to begin unambitiously, the simple language of propositional logic?

It seems to me that, for example, a straightforward interpretation of Gibbard [1986, 1990, 2003] does contain such package. Schroeder thinks not. So let's go through this.

### 2.1 Finding the problem with Gibbard's account

Building on Unwin [2001, 1999], Schroeder worries that Gibbard's account ultimately assumes something it needs to explain. What it assumes without explaining is the idea that there can be relations of inconsistency between pairs of attitudes which (i) differ in type, but which (ii) bear the same content. In Schroeder's jargon, what it assumes without explaining is "*B*-type inconsistency". This kind of consistency is meant to contrast with a more familiar kind of inconsistency, "*A*-type inconsistency". The idea of *A*-type inconsistency is the idea that there can be relations of inconsistency between pairs of attitudes which (i) are the same in type, but which (ii) bear inconsistent content. States of mind inconsistent in the *A*-type way are states whose inconsistency is typically *explained* (in part) by appeal to inconsistency between their contents. Belief states are supposed to furnish the paradigm examples of states susceptible to *A*-type inconsistency. E.g.:

- (1) Believing Paris is the capital of France.
- (2) Believing Paris is not the capital of France.

Perhaps states of intention supply further examples.

What are some examples of *B*-type inconsistency? There aren't any good examples, Schroeder says. That is why it is a problem that Gibbard's account calls for this kind of inconsistency. What is an example of the sort of *B*-type inconsistency Gibbard is forced to postulate? I found *Being For* less than direct on the matter. The main discussion of the issue occurs on pp. 52-4. There, Schroeder discusses what Gibbard must say about sentences such as:

- (3) Murder is wrong.

(4) Murder is not wrong.

Gibbard's account, at least on one direct reading, associates incompatible contents with these sentences. (In jargon I will review in more detail later, the sentences correspond to disjoint sets of centered world-hyperplan pairs.) So at first glance, the account seems to fit the *A*-type pattern, and the inconsistency between the states of mind characteristically expressed by these sentences can be explained in part by appeal to inconsistency between their contents. But, worries Schroeder,

... if we unpack it, what [Gibbard's account] is really saying is merely that 'murder is not wrong' must express a mental state that is inconsistent with all and only the hyperdecided mental states that 'murder is wrong' is not inconsistent with. And... that looks more like a list of the criteria that we hope the attitude expressed by 'murder is not wrong' will satisfy, in lieu of a concrete story about which mental state this actually is, and why it turns out to be inconsistent with the right other mental states. (52-3)

There is a good worry about explanatory adequacy in the vicinity of this passage, but it is not the advertised worry to the effect that Gibbard's account requires *B*-type inconsistency. I'd state the worry like this: Gibbard's style of explanation precludes the possibility of explaining the inconsistency between the states of mind characteristically expressed by (3) and (4) (etc.) in anything like the usual way—that is, by appeal, at least partly, to the inconsistency of their contents. This is because Gibbard assumes the reverse explanatory order, at least for his "plan-laden content": plan-laden content is literally modeled via sets of plan-laden states of mind, and inconsistency between plan-laden contents is ultimately supposed to be explained or explicated by a more basic inconsistency that exists between plan-laden mental states. (Or so anyway goes Gibbard's preferred gloss on his technical apparatus. Below I'll recommend an alternative gloss.) So Gibbard cannot see inconsistency between plan-laden states of mind as in part grounded by their incompatible contents. And that leaves an explanatory lacuna, one that appears to go unfilled. The worry is that without some deeper mooring of these inconsistency relations, the story starts to look dangerously just so.

If Schroeder is raising this sort of complaint, then I agree. We have a problem here. But this worry is not the worry that Gibbard's account

requires *B*-type inconsistency. The present worry would *remain* a worry even if it emerged that Gibbard’s account required only *A*-type inconsistency.<sup>1</sup>

So while I am sympathetic with Schroeder’s view that Gibbard’s account leaves something fundamental unexplained, I don’t see the rationale for his official diagnosis, that Gibbard’s view is problematic because it appeals in an unexplanatory way to *B*-type inconsistency. The passage quoted above does not bear on the question, as far as I can discern. Here is another key place where Schroeder discusses the issue:

... suppose you think that murdering is wrong. This state [according to Gibbard] is represented by the set of hyperplanners that 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 this disagreement can’t be mere *A*-type inconsistency. It has to include the *B*-type inconsistency that holds between disapproval and tolerance of murder. (54)

This passage seems to supply the desired argument. But the trouble is that the remarks after “But that means...” do not follow from what precedes.

To get clear here, let’s list the three states of mind that Schroeder distinguishes. Since the question partly at issue just is whether Gibbard’s model must deny that these three states belong to the same attitude type, let’s informally describe these states using the same attitude verb, so as not to beg questions in advance. What we want to find out is whether we are ultimately forced into talk of *B*-type inconsistency relations in order to characterize the

---

<sup>1</sup>To see this, consider a theorist who models the content of a belief state via sets of maximally opinionated belief states, and who appeals to primitive relations of inconsistency between the latter to ground consistency relations between belief contents. Such a theorist is blocked from the more routine explanatory order. She lacks a step of explanation her opponent possesses. She also lacks a nontrivial account of the content of maximally opinionated states. These are objectionable explanatory lacunae. This theorist obviously has some nontrivial explaining to do—and that is so even if she only recognizes inconsistencies of the *A*-type.

situation as Gibbard picture requires. Gibbard is quite comfortable allowing that ‘believes’ can be used to pick out the target plan-laden decisional states, so I’ll just follow him in this respect. Here are the three states, together with the corresponding properties a state of mind must have in order to count as instances of them, according to Gibbard:

- (a) Believing that murder is wrong.

Gibbard: all the hyperplans compatible with your decisional state fall within the set of hyperplans which forbid murdering

- (b) Believing that not murdering is wrong.

Gibbard: all the hyperplans compatible with your decisional state fall within the set of hyperplans which require murdering

- (c) Believing that both murdering and failing to murder are permissible.

Gibbard: all the hyperplans compatible with your decisional state fall within the set of hyperplans which allow-but-do-not-require murdering

Are we forced to postulate *B*-type inconsistency to model the states in Gibbard’s way? It seems not. Evidently an account with the structure of Gibbard’s has no problem saying that these three states of mind are all of the same type—they are decisional states of mind, or plan-laden belief states, or whatever—and yet they are pairwise incompatible, as reflected in (if not explained by) their mutually incompatible (disjoint) contents. I find nothing in what Schroeder writes that demonstrates that the account requires there to exist anything beyond *A*-inconsistency.

Of course, one could press the worry that even if there is only *A*-inconsistency here, we are left in a bad way for explaining it. That would be a good worry. It is the worry we already mentioned above. The point, again, is that that worry is distinct from the worry that Gibbard somehow needs *B*-type inconsistency.

Can a worry about *B*-inconsistency nevertheless be made out? It is not obvious, because the notion of *B*-type inconsistency itself is perhaps less than clear. Schroeder writes:

...*B*-type inconsistency is not something that expressivists can take for granted, because there are no good examples of it. (48)

Is this so? It is not totally obvious to me how we are supposed to play this game. Do we take ordinary language as a guide? In that case, there are seem to be perfectly good examples of *B*-type inconsistency. For instance:

- (5) Doubting that Paris the capital of France.
- (6) Being certain that Paris the capital of France.

Of course, one might say that at the relevant level of abstraction, these states of mind really do belong to the same type—they are, as we could put it, doxastic states of mind. But that is a theoretical move. It takes us from some intuitive, informal examples to a model aimed at systematizing and explaining the examples. Perhaps Schroeder’s thought is that after we find reflective equilibrium between (i) intuitive examples and (ii) worked out models of the mental states behind such examples, it will be clear that on the best models, there is no *B*-type inconsistency. Perhaps so, but it is certainly not obvious. So it is less than clear that *B*-type inconsistency is as pernicious as advertised—though we should happily agree that if it exists, it should of course be made intelligible.

## 2.2 What has the problem to do with the possibility of compositional semantics?

Setting the worry about *B*-inconsistency to one side, let’s agree that Gibbard has a problem on his hands in virtue of the fact that he cannot explain incompatibility between plan-laden states of mind partly in terms of incompatibility between their contents. What does this problem have to do with compositional semantics? In particular, what has it to do with the possibility question, the question whether Gibbard’s account can cohere with a compositional semantics for some possible language?

Very little, it seems to me. To state the trouble with Gibbard’s account, one really doesn’t even need to mention anything about language, as one would if the issue were one about compositionality proper. The problem with Gibbard’s package is not that it is somehow incompatible with a compositional semantics. It is that disagreement between plan-laden states of mind isn’t explained in a satisfying way. It is that more needs to be said there. If this raises any problem about semantics, it is a foundational one. It is a metasemantic problem. It is not a failure of descriptive semantics, as it would be if the account exhibited failures of compositionality.<sup>2</sup> There is a problem about disagreement, but there is no negation problem.

---

<sup>2</sup>Gibbard’s problem can be seen as one about showing that his account is compatible with a plausible, or even intelligible, theory of content. If one conflated the notions of *content* and *compositional semantic value*, one might see this problem as identical to one about compositional semantics. But one should not conflate these notions. They correspond to importantly different job descriptions. See Lewis [1980] for further discussion.

Separate from all this is the question whether Gibbard's semantic apparatus can be empirically motivated. I will return to that important issue below.

### 2.3 Explaining with hyperplans

Let's now directly confront the explanatory lacuna in Gibbard's account. How should the expressivist sympathetic to a broadly Gibbardian outlook respond? Three options come to mind.

1. Try to give a substantive, explanatory account of inconsistency between plan-laden/normative states of mind, not relying in the process on an antecedently grounded notion of inconsistency between the contents of the states.
2. Try to show that, despite appearances, explanations of inconsistency between mental states partly in terms of an antecedently explained notion of content are less explanatory than they superficially appear—so that the Gibbardian expressivist really doesn't have the explanatory disadvantage he appears at first to have.
3. Find an expressivist-compatible way to maintain the standard explanatory order, so that inconsistency between plan-laden/normative states can be explained partly in terms of inconsistency between their contents.

It appears Gibbard himself elects for some mix of options 1 and 2.<sup>3</sup>

I myself think it was ill-advised in the first place to attempt to explain inconsistencies in plan-laden content by appeal to brute disagreement between plan-laden states of mind. I don't deny this idea might be worked out somehow, but the path looks dark to me. I would sooner see where option 3 might get us.

Here again Schroeder and I evidently agree: the version of expressivism he explores in chapter 5 and beyond takes an approach in the spirit of the third option. We agree, too, that the expressivist account Schroeder ultimately goes on to construct is unsatisfying. But I want to consider now an alternative way of pursuing the third option. It employs Gibbard's formal framework and it embraces his talk of planning states, but it allows the conventional direction of explanation for understanding disagreement between such states.

---

<sup>3</sup>Judging from Gibbard (ms) 'Schroeder on Expressivism'.

To explain, let’s review Gibbard’s modeling proposal in greater detail. The proposal is abstract and applies directly only to idealized agents, but it provides enough structure to illuminate the target distinction between factual belief and states of decision. It begins with the model of belief given by Lewis [1979], in which a state of belief is represented as a set of centered worlds, intuitively the worlds not ruled out by what is believed. Gibbard adopts this much structure to model what it is to have a factual view, a view about how the world is (or about where one is within it). He then adds further structure to model what it is to have normative view (a view about what to do, a plan).

The additional structure Gibbard adds involves a technical notion, the hyperplan. As a single centered world might be used to characterize a state of mind completely opinionated concerning every matter of fact, so a hyperplan can be used to characterize a state of mind completely opinionated about what to do in any situation. A hyperplan is a maximal contingency plan: it

... covers any occasion for choice one might conceivably be in, and for each alternative open on such an occasion, to adopt the plan involves either rejecting the alternative or rejecting rejecting it. In other worlds, the plan either forbids an alternative or permits it. (56)

One can understand the word “open” here *externally*, as a matter of what is really open to the relevant agent in the relevant situation (whether or not the agent in that situation realizes it), or *internally*, as a matter of what the agent merely takes to be open in the situation.<sup>4</sup> Gibbard intends the latter. Given two centered worlds corresponding to distinct agents in epistemically indiscernible situations, a single hyperplan is meant to issue exactly the same verdict about what is permissible for the agents in these situations.<sup>5</sup>

In Gibbard’s formal development, hyperplan are taken as primitive in the way that possible worlds are taken as primitive in a possible worlds semantics. This primitivism about hyperplans is understandable, given some of the intended analogies between two notions, but it unfortunately obscures the important fact that hyperplans can easily be modeled entirely in terms of possibilities. It is clarifying to build hyperplans out of possibilities explicitly,

---

<sup>4</sup>To mention just two of many interpretive possibilities.

<sup>5</sup>Schroeder writes at one point that Gibbard has a problem because “he assumes that hyperplanners are always decided either to do *A* or to not do *A*, for any action *A*” (53). This claim is false: Gibbard makes no such assumption.



so I will do so here. As follows: a hyperplan is a function which takes a set of epistemically indiscernible centered worlds (a set which reflects a possible doxastic situation *vis-a-vis* what the world is like) to a set of centered worlds (where this set reflects the class of outcomes which are permissible to realize according to the hyperplan, given the evidential situation).

Now extending the Lewisian picture, the content of a plan-laden belief state is a set of centered-world, hyperplan pairs. Note that from the point of view our formal model, hyperplans are *not* identified with some idealized (hyperdecided) mental states; rather they are just certain functions on possibilia.

What now remains to be done, from the point of view of explanation? What remains to be done is just the sort of thing that would remain to be done if we had only the pure Lewisian picture of contents as sets of centered worlds. It remains to show how this model of what content is can perform the explanatory work associated with the notion of content. It would remain to show that we have here a plausible realizer for the content role.

What is the content role? This is a question of terminology, but how best to answer it is contested within the philosophy of content. Different ways of answering it reflect different conceptions of what the sensible explanatory priorities are. Fortunately, defending an answer to this question at any length seems unnecessary in the present dialectical context. A better move would be to sketch in broad strokes how the Lewisian picture of content is meant to be explanatory, and then ask whether a story along such lines can be extended in some intelligible way to the plan-laden upgrade. That would at least give some indication of how Gibbard's model compares to the descriptivist model it evolves from. Let us turn to this task.

Lewis, like most theorists, takes it that the contents of mental states are supposed to be (at least) causal-explanatory properties of them *vis-a-vis* action. Particular hypotheses about the belief and desire contents of an agent generate *ceteris paribus* predictions about how the agent will be disposed to act in various circumstances. So Lewis needs to indicate how centered worlds content carries a load in such explanations. Much of this task, as conceived by Lewis, is a matter of specifying how the functional interconnections between belief, desire, and action are sensitive to belief and desire content; and a significant part of that task is completed, for Lewis, by a banal hypothesis roughly along the following lines:<sup>6</sup>

If agent  $A$  is in a belief state with the centered content  $B$ , and in a desire state with content  $D$ , then  $A$  is disposed to act in ways

---

<sup>6</sup>See [Lewis \[1994\]](#), for instance.

that would tend to bring it about that he is located within  $D$ , were it the case that he occupied a centered world within  $B$ .<sup>7</sup>

This imposes a nontrivial necessary condition on belief and desire contents, and it makes intelligible how classifying states of mind via their centered contents can indicate something highly nontrivial about how agents in such states of mind are disposed to act. This does much to moor the abstract modeling proposal.<sup>8</sup>

Now if we could offer something analogous for the Gibbardian formal model, I think we would allay, to a substantial degree, the worry that modeling in terms of hyperplans is not at bottom explanatory. And it seems plausible that this can be done. I'll have to settle for a crude initial foray into the project of working this out. The suggestion will just be that this general way of approaching the issue has promise and merits investigation.

Rough idea: moving beyond the familiar belief-desire framework assumed by Lewis, we suppose that rational action centrally involves also states of planning, or (if you like) intention. We can expect the detailed functional interconnections between these three states to be elaborate, but it seems safe to hypothesize at least something like this: that agents are disposed to act in ways which would conform with their plans, in centered worlds with respect to which their belief content is true. Where such plans leave several options open, we add further that agents tend to elect those options which would serve best satisfy their desires. Finally, where agents find themselves in unplanned-for situations, we appeal only to belief and desire. That is, we say that such agents will be disposed to act in ways that would tend to satisfy their desires, in centered worlds where their beliefs are true—leaving intention out of it.

This goes some nontrivial way, I think, towards mooring plan-laden content. It is moreover difficult to discern any deep difference between this gen-

---

<sup>7</sup>True, desire should give way to preference, and belief to something like credence; and in that case we could instead frame the hypothesis as one about maximizing utility, or something thereabouts. But since we are familiar with such elaborations, let's just run with a terribly oversimplified picture.

<sup>8</sup>Of course, it remains to articulate the functional connections between belief, desire, and other states of mind, such as imagination. And of course, such functional interconnections do not suffice to uniquely determine belief and desire contents. To complete the task of determining these contents, Lewis adds an additional constraint on belief content, namely that the contents be sufficiently eligible. Externalists typically appeal instead to some additional 'wide' functional structure. These further issues strike me as largely orthogonal—the kind of expressivist I am now describing can accept any of the usual options here—so I set them aside.

eral way of mooring plan-laden content and Lewis’s general way of grounding his centered worlds content.

### 3 Empirical adequacy

The preceding has largely been about whether a certain model of content can be supplied with an adequate foundation. The issues can be framed and pursued largely in abstraction from theses about natural language. But of course, our expressivist wants to say, beyond all the preceding, that certain sentences of natural language serve to express plan-laden contents. This does not obviously constitute a thesis of semantics proper, but it does seem to place constraints on semantics. On the most straightforward development, it requires that the semantic values of sentences of natural language be the sorts of thing which can, relative to context, determine functions from world-hyperplan pairs to truth-values, or something equivalent.

It is perfectly possible for there to be such semantic values. But is there empirical motivation for the thesis that sentences of natural language have such semantic values?

It is striking, and I think dismaying, that Gibbard never attempts to put hyperplans to nontrivial semantic work: for instance, by articulating the compositional semantics of deontic modals, or the attitude verb ‘decides’, etc., by using hyperplans in the metalanguage. This is the sort of thing that would be required to really motivate the added structure from semantics-internal considerations.

Nevertheless, there is some empirical plausibility to the thesis. [Kratzer \[1991\]](#)’s account of deontic modals, with routine adjustments, can be interpreted as assuming hyperplans. Indeed, some recent work on deontic modals and conditionals in the broadly Kratzerian vein already contains objects with exactly the structure of hyperplans as formalized above (see the ‘deontic selection function’ of [Kolodny and MacFarlane \[2010\]](#); see also [Klinedinst and Rothschild \[2011\]](#)).

Anyway, even if it were established that the semantic values of sentences of natural language are in fact the sorts of thing which can, relative to context, determine nontrivial functions from world-hyperplan pairs to truth-values, the expressivist still has the burden of fleshing out and defending the idea that such functions are the contents of unembedded sentences of natural language. It seems to me this idea is best developed as a thesis in formal pragmatics—specifically, as a thesis about the characteristic sort of update to the conversational common ground a normative or plan-

expressing sentence serves to trigger. To say more: on what seems to me the most promising direction for development, the expressivist's idea is that the common ground, beyond reflecting what is mutually presupposed about matters of fact, admits of something like plan-like structure, and that some sentences—the normative or plan-laden kind—serve particularly to manipulate this structure, ruling admissible plans in and out. This, anyway, would flesh out what is meant to be distinctive about expressing plan-laden states of mind.

So developed, this kind of expressivism would best be understood as a pragmatic thesis about natural language—one which largely remains undeveloped and undefended, to my knowledge. Is it worth trying to develop? Well, allied forms of expressivism are viable,<sup>9</sup> so there is some reason to think that exploration here may yield fruit.

## References

- Allan Gibbard. *Thinking How to Live*. Harvard University Press, Cambridge, 2003. [2](#)
- Allan Gibbard. *Wise Choices, Apt Feelings*. Harvard University Press, Cambridge, 1990. [2](#)
- Allan Gibbard. An expressivistic theory of normative discourse. *Ethics*, 96(3):472–85, 1986. [2](#)
- Nathan Klinedinst and Daniel Rothschild. Connectives without truth tables. *Natural Language Semantics*, forthcoming, 2011. [11](#)
- Niko Kolodny and John MacFarlane. Ifs and oughts. *Journal of Philosophy*, CVII(3):115–43, March 2010. [11](#)
- Angelika Kratzer. Modality. In A. von Stechow and D. Wunderlich, editors, *Semantics: An International Handbook of Contemporary Research*, pages 639–50. de Gruyter, 1991. [11](#)
- David K. Lewis. Index, context, and content. In S. Kanger and S. Ohman, editors, *Philosophy and Grammar*, pages 79–100. D. Reidel, Holland, 1980. [6](#)
- David K. Lewis. Attitudes de dicto and de se. *Philosophical Review*, 88(4):513–43, 1979. [8](#)
- David K. Lewis. Reduction of mind. In S. Guttenplan, editor, *A Companion to the Philosophy of Mind*. Blackwell, 1994. [9](#)
- N. Unwin. Quasi-realism, negation and the frege-geach problem. *The Philosophical Quarterly*, 49(196):337–352, 1999. [2](#)

---

<sup>9</sup>Or so I've argued about a view I've called 'credal expressivism' in several papers.

N. Unwin. Norms and negation: A problem for gibbard's logic. *The Philosophical Quarterly*, 51(202):60–75, 2001. [2](#)