Draft of a paper forthcoming in Burgess and Sherman (eds), *Metasemantics: New Essays on the Foundations of Meaning*.
Please do not quote without permission.
Version 3.2. Last revision: 09.22.2013 at 00:32:10 -0400

SEMANTIC HERMENEUTICS

Alejandro Pérez Carballo
University of Massachusetts, Amherst
apc@umass.edu

Expressivism in metaethics (Blackburn 1998; Gibbard 1990, 2003, *inter alia*), as I will understand it, is a conjunction of two claims:

(E1) Moral thought is non-representational.

(E2) The meaning of a moral sentence is a function of the role it plays as a device for expressing moral thought.

My starting question is about the consequences of expressivism for the project of compositional semantics for natural languages. More specifically, I want to ask whether expressivism is compatible with a standard semantics for English. (Let me postpone for a moment the question of what a 'standard' semantics for English is.)

Contemporary orthodoxy presupposes:

> INCOMPATIBILISM: Expressivism in metaethics is incompatible with a standard semantics for natural language.

Indeed, many take the development of a semantics that is compatible with expressivism's core theses—the 'semantic program' of expressivism—to be the most important item on the expressivist agenda.

Here, I want to argue that expressivism is compatible with standard semantics. On the view that will emerge, expressivism answers a particular kind of metasemantic question, and not a question in semantics proper. What's more, the expressivist's answer to that metasemantic question is compatible with standard semantics. Or so I will argue.

Beyond providing a less theoretically costly way of thinking about expressivism, my goal is to highlight a cluster of metasemantic questions that have received little attention in the literature. In doing so, I hope to shed light on the role of formal semantics for answering questions in metaphysics and the philosophy of mind.

1

Before moving on, I want to spell out my initial characterization of expressivism, in terms of (E1) and (E2), in more detail. My goal here is not to defend either (E1) or (E2), but rather to offer some clarification of each of them.

Start with (E1): that moral thought is non-representational. To say that state of mind is representational is in part to say something about its functional role. As I will understand it, to say that a state of mind represents that its environment is in a given state is in part to say that its function involves systematically responding to the agent's environment being in that state. Thus understood, to deny that moral thought is representational is to deny that in order to understand the function of moral thought we need to appeal to putative moral facts to which our mental states aim to systematically respond.[1]

Note that (E1) does not imply anything about how our moral attitudes are implemented in our brains. It is compatible with (E1) that moral thinking involves manipulating sentence-like objects in a 'language of thought'.[2] What is ruled out by (E1) is that in order to understand the functioning of such sentence-like objects, we need to understand them as co-varying with features of the environment.[3]

Note also that (E1) is not supposed to be in tension with our commonsense view of ourselves as moral agents. The notion of representation that is relevant for understanding (E1) is a term of art. There may well be a pretheoretic understanding of 'representation' such that (E1) is compatible with

[1] Cf. Gibbard 1990, pp. 107ff. The issues here are complex, and I cannot do justice to them. See e.g. Dreier 2004, §VI for a related attempt at cashing out (E1) in the way I suggest; for critical discussion, see Chrisman 2008, esp. §III. Note that, on this intepretation, (E1) is something that error-theorists would deny. An error-theorist would presumably agree that what makes the belief that stealing is wrong have the content that it does is something about its functional role—it is the kind of state that, in normal circumstances, would indicate that stealing *is* wrong. For the relevant set of beliefs, the error-theorist claims, the 'normal' circumstances never actually obtain. Things are trickier when it comes to fictionalists who are not error-theorists, like so-called 'hermeneutic' fictionalists (no relation) in the philosophy of mathematics (cf. Kalderon 2005b, for an example of such a view applied to moral thought and talk). It is beyond the scope of this paper, however, to tackle the difficult interpretive question of how expressivist and fictionalist views differ from one another. For contrasting views on the matter, see e.g. Blackburn 2005; Lewis 2005. See also Eklund 2011 for a helpful guide to the terrain.

[2] If there is a non-relational understanding of 'representation' (cf. Chomsky 1995, p. 53), then (E1) is compatible with moral thought being representational *in that sense*.

[3] I want to remain neutral on what kind of *systematic response* is sufficient for a state of mind to count as representational. Some cognitive scientists explicitly stipulate that representational relations involve *causal* co-variation between states of the brain (say) and the enviroment (see e.g. Gallistel and King 2009, pp. 55f), but I leave it open whether other sorts of systematic co-variation could suffice.

moral thought counting as representational in *that* sense.[4] That said, I doubt ordinary usage puts much pressure one way or another. It may offend common sense to say that we have no moral beliefs. But I would be surprised if anyone were to insist that it is a 'Moorean fact' that in thinking that torture is wrong I represent the world as being a certain way.

Turn now to (E2): the meaning of a moral sentence is a function of the role it plays as a device for expressing mental states. To a first and rough approximation, this is nothing more than the claim that facts about the meaning of public language sentences are ultimately reducible to, or at any rate obtain in virtue of, facts about the mental states of users of the language—what, following Schroeder (2008b), we can call *mentalism*.[5] Thus understood, this is an non-trivial commitment of expressivism, and some of expressivism's most outspoken critics have explicitly questioned its plausibility. But mentalism is the majority view among philosophers of mind and language—it is a thesis with a rather impressive pedigree, and can be motivated and defended without having to take a stand on whether expressivism is the correct metaethical view.[6]

To some extent, (E2) goes beyond mentalism. It states not only that the meaning of sentences in a public language are somehow determined by the mental states of the speakers of that language, but that such determination goes via the 'expression' relation. So what is it for a sentence to play a role 'as a device for expressing mental states'?

This is a big question. For our current purposes, however, it will suffice to get a sense of what possible answers look like. I will briefly consider two.

The first, discussed by Schroeder (2008b), is that for a sentence to express a mental state is for it to be *semantically correct* to assert that sentence only when one is in that mental state (of course, for this to be of much help we need a story about what 'semantic correctness' amounts to). The sentence 'Bill was tortured' expresses the belief that Bill was tortured just in case there are

[4] E.g.: '[I] was thinking of writing a book representing moral principles via the use of empathy.' (http://able2know.org/topic/110960-1.)

[5] Cf. also Speaks (2011). The usage of this term is importantly different from Quine's. Quine uses 'mentalism' as a label for the view that meaning facts do not supervene on facts about verbal behavior. But facts about linguistic meaning could fail to supervene on speaker's dispositions to *verbal* behavior (i.e. mentalism *in Quine's sense* could be true) without meaning facts obtaining in virtue of facts about the mental states of users of the language. To confuse things further, Quine sometimes seems to use 'mentalism' as a name of the Lockean thesis that the meaning of words are 'ideas' in the mind of language users (cf. Quine 1964, p. 74f). This is certainly not part of mentalism as I understand it (nor is it equivalent to the denial of the supervenience of meaning facts on verbal behavior).

[6] See e.g. Davis 2002; Grice 1957, 1969; Lewis 1975; Schiffer 1982; Stalnaker 1984. Some notable exceptions include Davidson 1974 and Dummett 1991b.

linguistic rules in place for English whereby it is semantically correct to assert 'Bill was tortured' just in case one believes that Bill was tortured.

The second answer is a generalization of the picture sketched by Lewis (1975).[7] The general idea is that utterances are part of a practice whereby we intend to bring about changes in our audience's mental states, and that a semantic theory aims to partially characterize the role each sentence plays in such a practice. On one way of implementing this idea, sentences play the role they do in virtue of conventional associations between them (e.g 'Lucy likes Brahms') and certain mental states (e.g. the belief that Lucy likes Brahms). A sentence *S* in English expresses a mental state *M* just in case there is a convention in place whereby a speaker of English will only utter *S* when she is in *M*, and a hearer will come to be in *M* upon hearing an utterance of *S*. There are wrinkles that need to be ironed out, to be sure.[8] But I trust the general idea is clear enough for our purposes.[9]

Before moving on, I should note that (E2) does not directly tell us anything about what kind of things *meanings* are. In particular, it does not follow from (E2) that the meaning of a sentence is a mental state.[10] The claim is that (part of) what it takes for a sentence to mean what it does crucially depends on the role that sentence plays as a device for expressing mental states.[11]

2

Once we understand (E1) and (E2) along these lines, it is not hard to see why one might think that INCOMPATIBILISM is true. For if expressivism is true, the mental states that get expressed by moral utterances are non-representational. And the objects that a 'standard' semantic theory assigns to declarative English sentences are *truth-conditions*, or more specifically, sets of possible worlds. But it is hard to see how non-representational states—states which do not seem to have truth-conditions in the first place, since they do not represent the world

[7] Lewis himself gestures towards a somewhat similar generalization—see Lewis 1975, p. 171f.

[8] For framework conducive to doing so, see Stalnaker 1978. Cf. also Stalnaker 2002 on a more explicit characterization of the notion of presupposition and Yalcin 2007, p. 1007f on 'conversational tone'.

[9] A third srategy, in terms of a notion like *speaker meaning*, could be deployed here. But two is enough. At least for us, here.

[10] *Pace* Schroeder 2008a, ch. 2, I deny that an expressivist semantics needs to assign mental states as semantic values to well-formed sentences.

[11] I will slide between the version of (E2) as formulated in the text—restricted, that is, to moral language—and the fully general version, on which any meaningful sentence gets to mean what it does because of the role it plays as a device for expressing mental states. On this, I am in complete agreement with Schroeder (2008a, p. 22ff): the expressivist's commitment to mentalism has to be combined with a commitment to mentalism accross the board.

as being a certain way—could give rise to a view on which the meaning of moral sentances are sets of possible worlds.

Indeed, the received view seems to be that INCOMPATIBILISM is true. For example, Mark Schroeder writes (2008a, p. xi, xiii):

> Expressivism is a hypothesis about the semantics of natural languages. […] [T]he project of understanding how to construct an expressivist semantics is particularly pressing, both if expressivist views in any area of philosophy are to be taken at all seriously, and if we are to understand them well enough to see why they are false.

Similarly, Ralph Wedgwood takes it as a given that 'the fundamental explanation of the meaning of normative statements […] takes the form of a purely psychologistic *semantics* for normative statements, not a truth-conditional semantics'. (2007, p. 5; emphasis added)

More generally, talk of 'expressivist semantics' has gained wide currency as a name for a project—associated with Simon Blackburn and Allan Gibbard—of giving an alternative to standard textbook semantics for natural languages.[12] Expressivists and their critics thus seem to agree that in order to make expressivism a serious contender in metaethics, some hard semantic work needs to be done.

3

Be that as it may, expressivists should want to reject INCOMPATIBILISM. Giving an alternative to standard semantic theory is hard work. More importantly, INCOMPATIBILISM is in tension with some broader, and rather plausible, methodological principles that many expressivists explicitly accept. Let me mention just two.

First, expressivists have typically been sympathetic to some form of deflationism about truth (e.g. Gibbard 2003, p. x). Now suppose, as I think we must, that standard semantic theory is compatible with deflationism about truth.[13] In other words, suppose that we can accept the deliverances of contemporary semantic theory without giving up on a fully deflationary perspective on the notion of truth. Where then would the conflict arise between expressivism and standard semantic theory? True, standard semantic theory will assign to 'Torture is wrong' a set of truth-conditions. But on a deflationary

[12] E.g. Blackburn 1988. Cf. Rosen 1998 on Blackburn 1993. Cf. also Sinclair 2009, p. 142: "The most serious challenge facing any expressivist position is explaining how the distinctive features of the target discourse can be generated by *an underlying expressivist semantics*" (my emphasis).

[13] See Burgess 2011 for a detailed defense of this claim.

view, an expressivist is perfectly entitled to thinking that 'Torture is wrong' has truth-conditions.[14]

Second, in motivating their view, expressivists often appeal to a particular form of *naturalism*. The idea is hard to state precisely, for reasons that should be all too familiar, but very roughly: a philosophical theory should be compatible with the picture of the world that our best science has to offer.

Gibbard and Blackburn have both explicitly appealed to naturalism in motivating their own versions of meta-ethical expressivism. For example, Blackburn (1998, p. 48):

> The natural world is the world revealed by the senses, and described by the natural sciences: physics, chemistry, and notably biology, including evolutionary theory. However we think of it, ethics seems to fit badly into that world. […] To be a naturalist is to see human beings as frail complexes of perishable tissue, and so part of the natural order.[15]

Naturalists have long been skeptical of any philosophical view that conflicts with the deliverances of the natural sciences. Philosophy, on this view, should never proceed on the assumption that it has the epistemological high-ground with respect to the sciences.[16] If constraints internal to semantic theory tells us that 'Torture is wrong' has a certain set of worlds as its semantic value, then so be it.

Admittedly, naturalism does not entail that INCOMPATIBILISM must be false if expressivism is true. At the end of inquiry, semantic theory may take a shape radically different from any one we now know of. The issue of the compatibility of expressivism with *current* textbook semantics may turn out to be moot. And perhaps the questions that exercise expressivists in meta-ethics will, in the long run, morph into straightforward questions in empirical psychology. If we think that linguistics is a branch of psychology, or cognitive science, it may then turn out to be good methodology to have our compositional semantics heed the advice of our best theories of the mind. But as things stand, there are obvious tensions between naturalism and INCOMPATIBILISM.

[14] This question has been extensively discussed in the literature. See e.g. Boghossian 1990; Dreier 2004; Kraut 1993; O'Leary-Hawthorne and Price 1996; Williams 1999, 2010.

[15] Cf. also Allan Gibbard's remarks on 'the great successes of the broadly Galilean view of the world', in Björnsson and Bâve 2007.

[16] Although Quine was probably the most ardent proponent of this methodological stance, David Lewis' characterization is particularly memorable: "Mathematics is an established, ongoing concern. Philosophy is as shaky as can be. To reject mathematics for philosophical reasons would be absurd […] Even if we reject mathematics gently—explaining how it can be a most useful fiction[…]—we still reject it, and that's still absurd […] That's not an argument, I know. Rather, I'm moved to laughter at the thought of how presumptuous it would be to reject mathematics for philosophical reasons." (Lewis 1991, pp. 58–59)

The marriage is not an easy one. We should thus be suspicious of expressivism if it requires that we give up on textbook semantics.

4

Can expressivists reject INCOMPATIBILISM?

Some have recently suggested that we understand expressivism as a *metasemantic* thesis. The idea is that expressivists can do justice to their core commitments while holding on to standard semantic theory if they provide an alternative metasemantics for moral language. On this view, expressivist and descriptivists will agree on what is the right semantic theory for moral language. Their differences will only be reflected at the level of their metasemantic theory.

For example, Matthew Chrisman has argued that we should think of expressivists as agreeing with descriptivists on the question of what possible-world propositions get assigned as contents of English sentences, but disagreeing on the question of 'what it is in virtue of which particular kinds of words have the semantic contents that they do' (2012, p. 325).[17] Michael Ridge is also developing a version of expressivism that is meant to be be compatible with a truth-conditional approach to first-order semantics. What is supposed to be distinctively expressivist about his view is the explanation of why moral sentences have the truth-conditions that they do.[18]

On different grounds, Huw Price suggests that expressivism is best understood as a thesis about 'how there come to be descriptive contents, or thoughts, of particular kinds' (2004, p. 184). More specifically, Price thinks that expressivism is compatible with thinking of the meaning of moral sentences as given by the sentences truth-conditions. The key, he thinks, is to see that expressivism 'provide[s] a pragmatic account of how there come to be the kind of judgements whose contents may be specified' by a given assignment of truth-conditions (p. 186).[19]

[17] See, also, p. 327:"[V]iewing realism and expressivism as competitors to the possible world semantics conflates an issue in semantic theory with an issue in the foundational theory of meaning (or 'metasemantics'). By viewing realism and expressivism not as different views about the semantic contribution of 'ought' but as different views about why it is that this word has the semantic value that it has, I believe we begin to usefully reorient metaethical debate about the meaning of this term."

[18] See Ridge forthcoming. Another recent example is Jussi Suikkannen, who has claimed that the debate between expressivists and others is one about 'in virtue of what the predicate 'is wrong' has that particular semantic value.' ('Metaethics, Semantics, and Metasemantics', available at: http://peasoup.typepad.com/peasoup/2009/07/metaethics-semantics-and-metasemantics.html.)

[19] Price here relies on the distinction between modest and full-blooded theories of meaning, in the sense of Dummett 1975. His claim, as I understand it, is that expressivists can stick to a

There is something very natural about these suggestions. Indeed, they look like the only way of meeting two desiderata: first, that of giving a non-revisionary semantic theory; second, that of maintaining that there is something distinctive that expressivism has to offer for our theories of the *meaning* of moral language. After all, it seems that a theory of meaning should encompass nothing more than a semantic theory and a metasemantic theory. So if expressivism does not require giving up on standard semantic theory—if INCOMPATIBILISM is false—its only distinctive contribution to a theory of meaning has to be at the level of metasemantics.

At the same time, there is something puzzling about them. To fix ideas, consider the following simple example:

(1) Torture is wrong.

A fairly orthodox semantic theory would treat (1) as follows:[20]

(2) $[\![\textbf{Torture is wrong}]\!]^w = 1$ iff Torture is wrong at w,

or alternatively:

(3) $[\![\textbf{Torture is wrong}]\!] = \{w : \text{Torture is wrong in } w\}$.[21]

According to the meta-semantic construals of expressivism just mentioned, both the descriptivist and the expressivist will endorse (3) as the correct entry for (1). The disagreement will only turn up when they get to explaining *why* (3) is the correct entry for (1).[22]

---

standard semantic theory (say, one given in terms of truth-conditions) for the purposes of a modest theory of meaning. What is distinctive about expressivism as a theory of meaning will only become apparent once we move on to give a full-blooded theory of meaning.

But it is a mistake to think that what is needed to go from a modest to a full-blooded theory of meaning is an explanation of 'how there come to be the kind of judgments whose contents may be specifed' by a given assignment of truth-conditions. In my view, a full-blooded theory of meaning must first of all answer what I will call the *hermeneutic* question.

[20] Of course, much of the interest in such a theory comes from the fact that it provides a derivation of such an assignment in a compositional fashion. The bare-bones entry in (2) abstracts away from this and more. For our purposes, however, this toy version will do.

[21] For technical reasons that need not concern us here, textbook semantics for (2) assigns to 'Torture is wrong' the characteristic function of this set. For ease of exposition, I will talk as if, on the standard picture, all well-formed declarative sentences get assigned sets of possible worlds.

[22] Chrisman, Price, and Ridge may well disagree on the *kind* of explanation that they take expressivists to offer—is it a metaphysical explanation or a causal-historical one? I cannot tell. In fact, it is not obvious whether there is an 'orthodox' understanding of this question. Some of the classic texts in metasemantics appear to identify the metasemantic question with a historico-sociological one (as suggested by Kaplan 1989, p. 573f, as well as the title of Almog 1984). Others suggest it is the metaphysical one of what determines that a particular word has the semantic value that it has (e.g. Stalnaker 1997, p. 535).

But what is it for (3) to be the correct entry for (1)? It seems natural to suppose that, in assigning a set of worlds to (1) as its semantic value, (3) is taking a stance on whether an utterance (1) is representational: it represents the world as being a member of that set. And if an utterance of (1) is representational, it is hard to see why the belief that torture is wrong is not—in other words, it is hard to see how (E1) could be true.

One could insist that none of this follows from assigning a set of worlds to (1) as its semantic value. But the *descriptivist* thinks that it does. This indeed is the thought motivating the case for INCOMPATIBILISM. From the descriptivist's point of view, a metasemantic theory should explain how a string of symbols lile (1) could get to have the representational properties that it does: that is what she takes the explanandum to be. The expressivist cannot agree that this is what needs to be explained without giving up on the claim that moral thought is non-representational.[23]

Alternatively, one could insist that the explanandum is not what the descriptivist takes it to be. Rather, the explanandum is simply the fact that a sentence has a particular abstract object as its semantic value. But what kind of fact is that? What is it for an English sentence to have a semantic value? Presumably, whatever *having a semantic value* amounts to, what semantic value a sentence has will tell us something about the meaning properties of the sentences. The interesting explanatory question—the one that metasemantic theories typically aim to answer—is why words and sentences have the meaning properties that they do. It is only after we have answered the question of what an assignment of semantic values tell us about the meaning properties of the relevant sentences that the explanatory question can arise.[24]

We should understand the expressivist and the descriptivist as disagreeing over a metasemantic question. But it is important to be clear on what that question is. We need to distinguish the question, *what does an assignment of semantic value tell us about the meaning properties of a sentence of English?*

[23] One of the main motivations for (E1) is the lack of a good story of how creatures like ourselves could get to stand in representation relations with what moral facts would have to be like. If we grant that the sentence 'torture is wrong' has a representational content, we seem to undermine this motivation for (E1). Claims about the motivating character of moral belief may well suggest that moral thought is importantly different from non-moral thought, but not that moral thought cannot be representational at all.

[24] This needs to be qualified. Suppose you learn: 'bovino' (in Esperanto) means *cow*. You ask: what makes it the case that 'bovino' means *cow*? Without knowing what it is for 'bovino' to mean *cow*—just by knowing that 'bovino' in Esperanto means the same that 'cow' (in English) means—you can ask that question; perhaps you may even be satisfied with the following answer: because L. L. Zamenhof stipulated that 'bovino' was to mean the same as 'cow'. But the explanatory question—the real question that foundational theories of meaning aim to answer—has thus only been postponed. More on this below.

from the question, *in virtue of what does a given sentence of English has the meaning that it does?* The expressivist and the descriptivist can agree on the structure of the formal semantics: expressivism's commitments for the theory of meaning are largely metasemantic. Their fundamental disagreement, however, is over what that semantic theory tells us about the meaning properties of moral sentences. They may have different theories about why a given sentence has the meaning properties that it does. But they disagree on what they take that explanandum to be.

5

At a high level of abstraction, the distinction I am after is one between

> What is it for theory T to be the correct theory of subject matter M?

and

> In virtue of what is theory T the correct theory of subject matter M?

The first asks what a given theory tells us about the world. The second asks for an explanation of why the world is the way the theory tells us it is.

But it is best to start with an example. Open a textbook on quantum physics and you are likely to find something like the following claim somewhere:

(4) The probability that a radium atom decays within a period of 1601 years is 1/2.

(Or, in the jargon: a radium atom has a half-life of 1601 years.)

Now:

(5) What is it for (4) to be true?

Or: what does (4) tell us about the world? Two potential (partial) answers:[25]

(6) a. That the author assigns a degree of belief of 1/2 to a given radium atom decaying within a period of 1601 years.
b. That the frequency of radium atoms decaying within a period of 1601 years is 1/2.

[25] To be sure, the answers in (6) could be offered as answers to the explanatory question. For example, we might think that probability facts do not reduce to, nor can they be analysed in terms of, facts about degrees of belief, and nonetheless think that probability facts are grounded in, or explained in terms of, facts about degrees of belief. For present purposes, I only want to consider the thesis in (6) as candidate answers to (5).

These two answers are not equivalent. The author could well be mistaken about her degrees of belief. But it is in principle *possible* for her to find out what degree of belief she assigns to a radium atom decaying within a period of 1601 years. In contrast, it could well be that facts about frequency are beyond the author's epistemic reach. Indeed, if (6a) is the right way to understand (4), then in order to find out whether (4) is true we only need to have access to facts about the author's state of mind; if instead (6b) is the right way to understand (4), then we need access to much more than the author's state of mind in order to determine whether (4) is true.

Perhaps there is conclusive reason to think that neither (6a) not (6b) can be the right way to understand (6). It may be crazy to suppose that a physics student would have any interest in knowing the state of mind of the textbook's author. And it may be crazy to think that the probability facts that quantum mechanics talks about are simply facts about relative frequencies. Even so, each of (6b) and (6a) is an attempt at accounting for what (4) says about the world.[26]

In some sense, each of the answers in (6) are *interpretations* of (4). But talk of interpretation might lead to misunderstandings. It is part of the job of a theory to tell us how its theoretical terms are to be interpreted. And those interpretations had better have something to do with the theory's intended subject matter. This may sometimes require no additional work: it may be clear from the outset how each of the terms of the theory is to be understood. But if the theory introduces technical terms, perhaps governed by certain formal assumptions, the theory should include a specification of how to interpret those terms.

Sometimes, one can say how a theory is to be understood by attaching familiar meanings to the theoretical terms. Sometimes, however, this is not fully satisfactory. The familiar meanings in terms of which the theory is being explained may themselves not be well-understood, or they may be governed by conflicting assumptions (to say nothing of those cases in which the theoretical terms do not correspond to more familiar ones, where appealing to something like Ramsey sentences, as in Lewis 1970b, may be called for). In such cases, one can hope for a more illuminating account of how the theory is to be understood. And while it is part of a theory to offer the first kind of explanation of its terms, lack of a philosophically satisfying account of its subject matter need not be impede theoretical progress. Physical theories tell us much about space-time. But there are plenty of questions about the nature of

[26] To quote Bennett 2009, this is one of those instances where "the somewhat tendentious 'nothing over and above' locution is apt" (p. 47). The claim being made in the answer given in (6b), say, is that the fact that a radium atom has a half-life of 1601 years is 'nothing over and above' some fact about relative frequencies.

space and time that one might want answered that physics seems to have little to say about.[27]

Once we think of (5) that way—as a request for an illuminating account of the subject matter of (4)—it is tempting to identify the task of answering (5) with that of giving an *analysis* of the property of having probability 1/2 of decaying within 1601 years.[28] Thought of that way, it seems that there is an important distinction between the hermeneutic question and the explanatory question, in virtue of what does a radium atom have a half life of 1601 years? The latter question will presumably need to be answered by appealing to facts about radium. The former, however, may well not—answers like (6a) make no substantive appeal to any facts about what radium is.[29]

But this way of thinking about (5) is not quite right.[30] An additional complication arises out of the fact that statements like (4) are made against the backdrop of a particular formalism governing the technical notion of probability. There are two sources of constraints on a satisfactory account of what probabilities are. On the one hand, we have the ordinary notion of probability. On the other, we have the notion of probability characterized by particular mathematical functions, satisfying certain formal constraints. A statement like (4) is thus a bit of applied probability theory. The axioms governing the formal notion of probability can be seen as characterizing a functional role. To

[27] Cf. Stich 1992, §5: "Sometimes the relevant science will be pretty explicit about how it conceives of the item of interest. *The Handbook of Physics and Chemistry* will tell you all you want to know about gold, and then some. But in lots of other cases a science will use a concept quite successfully without providing a fully explicit or philosophically satisfying account of that concept. In those cases, philosophers of science often step in and try to make the notion in question more explicit." (p. 251)

[28] Or perhaps: a real definition of that property. This would involve reading (5) as a constitutive question, one asking about the 'essence' of the relevant fact. Alternatively, we could think of (5) as asking for a reduction of (4). However, talk of reduction is tricky. Thinking that there must be an answer to (5) should not involve thinking, to paraphrase Fodor (1987, p. 97), that if probabilities are real, they must really be something else. In so far as talk of reduction is entangled with some kind of eliminiativism, we should not think of (5) as a request for a reudction of (4). Yet another possibility, which I will set aside for present purposes, is to think of (5) as a request for an analysis of our *concept* of probability.

[29] The distinction can be nicely illustrated if we think of disjunctive claims. The fact that either Obama is the president of the United States in 2012 or Romney is the president of the United States in 2012 obtains *in virtue of* the fact that Obama is the president of the United States. But it does not seem at all obvious that for it to be the case that either Obama is the president of the United States or Romney is the president of the United States *just is* for Obama to be the president of the United States. For careful discussion of the relationship between ground and reduction, see Rosen 2010, §10. For a defense of the distinction between ground and essence, see Fine 2012, §11, as well as Rayo 2013, esp. §1.1 for more on the difference between explanatory questions and 'what it is' questions. Greenberg 2005, p. 304f, makes a related distinction, in discussing what he calls "different kinds of constitutive accounts, with different ambitions."

[30] Cf. Hájek 2012 on what interpretations of probability amount to.

apply the formal theory of probability, we need to point to some feature of the world that can be seen as playing that role. The kind of application in question here, however, is more heavily constrained. We want the realizer of the probability role to be able to play the role that probability plays in our ordinary lives ('probability is the very guide of life', after all), and as a result we wil need to specify the relevant functional role in ways that go beyond the particular axiomatization of probability in play (cf. Lewis 1970b).

I will call questions like (5), understood along these lines, *hermeneutic* questions. The hermeneutic question for (4) can be thought of as involving both a bit of interpretation of the formalism—attaching familiar interpretations to the primitives of the formal theory, so as to make true claims about the given subject matter—together with a bit of analysis in the above sense.

It is important to emphasize that to seek an answer to (5) is not to enshrine the ordinary, pre-theoretical notion of probability. Probability theory is a branch of mathematics, and as such it is not hostage to the idiosyncracies of our ordinary probability talk. And in order for such a theory to be fruitfully applied so as to understand the nature of stochastic processes, say, the theory does not need to neatly map onto our pre-theoretic notion of probability. It is up to the theorist to specify what properties of the relevant events are being explained and illuminated by this project. But in so far as the theorist's goal is to tell us something about the probabilities of coin tosses and what not, there had better be some connection between the properties of the events being investigated by a particular application and our pre-theoretic notion of probability.

The claim that a particular event has been assigned a numerical value for the purposes of modeling some of its properties is not by itself something that cries out for an explanation. It is only after we interpret that claim as one about some specific properties of the relevant event—after we answer the hermeneutic question—that we can raise the explanatory question.[31] Once we find a realizer for the probability role, we can ask what makes it the case that it plays the particular functional role it does. If we go for a subjectivist interpretation of probability we can ask, for example, what makes it the case that our credences obey the axioms of the probability calculus, or what makes it the case that our credences determine what sorts of bets we ought or ought not take. These may be hard questions to answer. But they are the kind of explanatory questions that arise only once we have settled on an answer to the hermeneutic question.

As should be clear by now, the explanatory project cannot be neatly separated from the hermeneutic one. It may be that we cannot engage in one of these projects in isolation from the other. Explanatory constraints will help

[31] Although see fn. 24.

shape what we take to be plausible answers to the hermeneutic question. And different explanatory strategies may be more or less attractive depending on what we take semantic facts to be. The best methodology here may be some form of reflective equilibrium. But whichever way we go, we should not lose sight of the fact that there are two different questions here, and it is important to keep them apart.

6

The hermeneutic question can be asked for attributions of semantic values, much as with attributions of probability. Much like we need an account of what it is for a probability claim like (4) to be true, we need an account of what it is for a particular lexical entry like

(3) $[\![\textbf{Torture is wrong}]\!] = \{w : \text{Torture is wrong in } w\}$.

to be correct. More generally, we need an account of what it is for a particular set to be the semantic value of a given sentence. Call *semantic hermeneutics* the project of answering the hermeneutic question for claims about semantic values of English sentences. Henceforth, I will restrict the term *semantic theory* to an assignment of semantic values to sentences in a language. I will reserve the term *theory of meaning* for a theory that includes a semantic theory together with an answer to the hermeneutic question for that theory, as well as an answer to the corresponding explanatory question.

The notion of semantic value, like the notion of probability, is a technical one, one that is governed by certain formal principles. Foremost of all, semantic values obey the principle of compositionality, and they combine in ways that follow a number of rules. A textbook semantics, like that in (Heim and Kratzer 1998), assigns to the syntactic constituents of a given sentence an object of some *type* or other (its semantic value). It then specifies the formal principles governing which types of semantic values are allowed to combine with one another, and how the semantic value of the complex expression (and its type) is a function of the semantic value of its simpler constituents.

For example, a textbook semantics will assign to the syntactic constituents of (1) the following semantic values:

(7) a. $[\![\textbf{Torture}]\!]^w$ = Torture.
b. $[\![\textbf{is wrong}]\!]^w = \lambda x.x$ is wrong at w.[32]

[32] I oversimplify. We may expect the lexical entry should tells us that the relevant reading of 'wrong' applies only to actions or action-types. And it may be that, at a deeper level of analysis, the right logical form of (1) involves quantification over events. But let me set such complica-

As a resut, the theory will yield (2) as the entry for 'Torture is wrong', as follows:

$$\begin{aligned}[\![\textbf{Torture is wrong}]\!] &= \{w : [\![\textbf{is wrong}]\!]^w([\![\textbf{Torture}]\!]^w) = 1\} \\ &= \{w : \text{ Torture is wrong at } w\}\end{aligned}$$

For our purposes, the details do not matter. While the resulting theory is more sophisticated, it comes down to something much like the familiar model-theoretic semantics for first-order logic. In each case, we have functions which take syntactic objects and assign an interpretation or semantic value—where abstract objects, typically set-theoretic constructions—to each of them in a way that obeys certain basic principles. The resulting theories yield assignments of semantic values to infinitely many sentences that are generated from an assignment of semantic values to a finite number of syntactic constituents.[33]

Now, assignments of semantic values to English sentences call for interpretation. We are told by a given semantic theory that a certain sentence gets assigned, by the semantic theory in question, some particular abstract object—in this case, a set of possible worlds. What does that assignment tell us about the properties of the relevant sentence that semantic theorizing set out to investigate?[34]

One straightforward answer would be: the assignment of semantic value to the sentence gives its meaning—in other words, the meaning of the sentence is given by the semantic value that the theory assigns to it. But this would not be a satisfactory answer. Our ordinary notion of meaning is notoriously messy. While it may be true that our semantic theories set out to investigate the meaning properties of English sentences, we need a more philosophically illuminating account of what those properties are. We need to know, say, what would count as evidence for or against a particular assignment of semantic value. This requires a careful specification o the properties of natural language sentences that semantics aims to investigate, and of the way in which we can read off claims about those properties from a given assignment of semantic values.[35]

tions aside.

[33] Roughly. The relationship between syntax and semantics is much less straightforward than this might suggest. For a sample of some of complex ways in which syntax and semantics interact with one another, see e.g. Higginbotham 1985.

[34] Cf. MacFarlane 2010, p. 83: "If formal semantics is to have anything to do with meaning (as opposed to being a rather ugly branch of algebra), its basic concepts must have significance beyond their structural role in the formal theory."

[35] Providing such an account is not a matter of giving an analysis of our ordinary concept of meaning (nor of whatever property our ordinary talk of meaning succeeds at picking out).

A more promising way of interpreting talk of semantic values would be to say that the assignment of a set of possible worlds to a given sentence is telling us something about the truth-conditions of that sentence. The way to understand (2) is as a formalized version of the following:

(8) 'Torture is wrong' is true at *w* if and only if torture is wrong at *w*.

We could then move from the claim that 'Torture is wrong' has a certain semantic value to the claim that 'Torture is wrong' *means that* Torture is wrong.

But this is only the beginning of the answer. We still need to hear what it is for 'Torture is wrong' to mean that Torture is wrong (or to be true at *w* iff torture is wrong at *w*). And unless we know more about what truth-conditions are, this way of understanding (2) does nothing to tell us what that meaning fact consists in.

Whether such an answer can be developed in more detail, the point stands: a compositional assignment of semantic values, or an assignment of truth-conditions, to sentences of English is at best an incomplete account of the meaning of the relevant expressions. We need an answer to the hermeneutic question in order to get a satisfactory theory of meaning out of any compositional semantic theory.

A very similar point was made by Michael Dummett in discussing truth-theoretic approaches to meaning.[36] The observation has been much discussed before, so I will be brief.

Start out by making the distinctively Dummettian assumption that 'a theory of meaning is a theory of understanding'. A complete theory of meaning for a given language must therefore explain what a speaker has to know in order to understand that language—to know the meaning of sentences in that language. But a theory consisting of a series of axioms from which we can derive T-biconditionals for each sentence in a language does not do much to explain what someone needs to know in order to know the language. Here is Dummett, in full (1975, p. 8ff):

[…] if we are asked whether the M-sentence "'La terra si muove' means that the Earth

Rather, it is a matter of giving an adequate account of the object of study of semantic theory.

[36] Dummett's observation was originally presented (in the lecture that is the basis of Dummett 1975) as an objection to Davidson's truth-theoretic account of meaning. In his terminology, he took Davidson's theory of meaning to be a modest one, and thus to be unable to give a full account of understanding—what, according to Dummett, was an essential part of what a theory of meaning should do. Later on, he came to the conclusion that his objection was based on a misreading of Davidson's view. As he put it in the appendix to the published version of the lecture in Dummett 1975: "The conclusion to which I am driven is that it is, after all, a mistake to view a Davidsonian theory of meaning as a modest one in any sense" (p. 26). Cf. also Dummett 1991a, p. 107ff

moves" expresses what someone has to know in order to know what the Italian sentence "La terra si muove" means, we can hardly do other than answer affirmatively: to know that "La terra si muove" means that the Earth moves is just to know what "La terra si muove" means, for that is precisely what it does mean. If, on the other hand, we are asked whether an adequate account of what a knowledge of the meaning of "La terra si muove" consists in is given by saying that one must know what is stated by the relevant M-sentence, then, equally, we are impelled to answer negatively: for the M-sentence, taken by itself, is, though by no means uninformative, signally unexplanatory. […] The simplest way we have to state its unexplanatory character is by observing that we have so far found no independent characterization of what more someone who knows that the M-sentence is true must know in order to know the proposition it expresses, save that he must know what "The Earth moves" means: knowledge of that proposition cannot, therefore, play any part in an account of that in which an understanding of that sentence consists.

Now, if we had an account of what it is to know the meanings of the primitive terms in the language, and of how these bits of knowledge combine to yield knowledge of the meaning of more complex expressions, we could use a Tarski-style theory of truth for a language to give an account of what knowledge of that language consists in. The problem is that a theory of truth does not, by itself, tell us what knowing that 'Earth' means Earth amounts to.

A similar complaint can be formulated even if we do not assume that a theory of meaning is a theory of understanding. Assume instead that a complete theory of meaning must be 'a complete theory of how the language functions as a language' (Dummett 1975, p. 2). The complaint against Tarski-style theories of truth would then take the following form: A theory that simply tells us that 'Torture is wrong' is true iff torture is wrong, and so on for any other sentence in the languge, will not be much of an account of how the language functions 'as a language'.

It is worth emphasizing that is not an objection to appealing to Tarski-style theories of truth in giving a theory of meaning. It is just to say appealing to such a theory is not enough. In providing a theory of truth for English, a theory of meaning has offered a semantic theory for English—it has specified semantic values for each English sentence. But that cannot be the end of the story: we need to be told how to interpret that semantic theory. We need to be told what it is for a given theory of truth to be correct.

This observation applies, *mutatis mutandids*, to theories of meaning that appeal to compositional assignments of truth-conditions to sentences in English. Gilbert Harman famously made this point, in arguing for some form of conceptual role semantics (1974, p. 196):

[…] there is a sense in which a theory that would explain meaning in terms of truth conditions would be open to Lewis's objection to Katz and Postal's theory of semantic

markers. Lewis says [1970a], you will recall, 'But we can know the Markerese translation of an English sentence without knowing the first thing about the meaning of an English sentence: namely the conditions under which it would be true'. Similarly, there is a sense in which we can know the truth conditions of an English sentence without knowing the first thing about the meaning of the English sentence. To borrow David Wiggins's (1972) example, we might know that the sentence 'All mimsy were the borogroves' is true if and only if all mimsy were the borogroves. However, in knowing this we would not know the first thing about the meaning of the sentence, 'All mimsy were the borogroves'.

The point here also does not depend on enshrining some pre-theoretic notion of meaning—we need not assume that ordinary practice sets the standard against which a semantic theory must be measured. The point is rather that, if 'All mimsy were the borogroves' has interesting semantics properties, we will know little about what those properties are just by being told that the semantic value of 'All mimsy were the borogroves' is the set of worlds in which all mimsy were the borogroves.

Harman takes this observation to be a reason for developing an alternative to truth-conditional semantics. But if I am right, we can grant Harman's observation without giving up on truth-conditional semantics. Once we acknowledge that a semantic theory is only part of a full theory of meaning—which will also include answers to the hermeneutic question and the explanatory question—Harman's observation is harmless. A theory of meaning that appeals to a particular semantic theory should answer the hermeneutic question for that theory. The two projects—the project of developing a semantic theory and that of answering the hermeneutic question for that theory—are no doubt related. But they are conceptually distinct projects, and it pays to keep them apart.[37]

[37] Something close to the distinction I am making here has been made before. For example, in discussing Davidsonian theories of meaning, Richard Heck draws a distinction between two different projects in Davidsonian theories of meaning (Heck 2007, p. 538): "The first is the semantic project of actually developing a theory of truth for a natural language, that is, a theory sufficient to yield theorems stating the semantic properties of all expressions of English (and to systematize that collection of facts by deriving those concerning complex expressions from those about their simpler parts). The second is the meta-semantic project of answering the question what it is for English expressions to mean what they do." However, Heck explicitly identifies 'the question what it is for expressions to mean what they do', with the question 'what determines what they mean, in a metaphysical sense' (cf. p. 533)—to that extent, the distinction he is drawing is not quite the same as mine. Cf. also Williams 1999, p. 553: "We can now see more clearly why, when thinking about Davidson, we must be careful not to conflate the two uses of the term 'theory of meaning'. Using the term narrowly, as Davidson himself often does, we refer to some axiomatic theory, a recursive device for specifying the meaning of every sentence of a given language or, more precisely, of the current idiolect of a particular speaker. No such particular theory constitutes an account of what meaning *consists in*, however [...]".

It also pays to distinguish the hermeneutic question from the explanatory question that Chrisman and Price focus on. Suppose we say, as a partial answer to the hermeneutic question for a truth-conditional semantics:

(9) Part of what it is for the meaning of the sentence 'Bill was tortured', in English, to be given by the set of worlds in which Bill was tortured is for 'Bill' to stand for Bill (the man) himself.

We can now ask the explanatory question: *in virtue of what* does 'Bill' stand for Bill himself? The story in (9) does not answer it. Indeed, it is a familiar point from discussions of theories of reference since at least (Kripke 1980) that we can have very different explanations of the fact that 'Bill' stands for Bill himself. These explanations—say, one in terms of an initial baptism and an appeal to some causal chain or other—are not plausibly construed as explanations of the fact that Bill is the semantic value of 'Bill'. They are explanations of that claim only if we interpret it along the lines of (9). We could instead think that what it is for 'Bill' to mean Bill is for the public language term 'Bill' to be associated a specific concept or mental representation. And if that were how we understood what it is for 'Bill' to mean Bill, we would not be satisfied with an answer that appeals to some causal link between utterances of 'Bill' and Bill himself. For not any such causal link would involve concepts or mental representations at all.

Once we recognize that there are two different kinds of metasemantic questions—the hermeneutic question and the explanatory question—we can ask whether we can reconceive of expressivist's commitments in the theory of meaning as largely metasemantic. But it is in answering the hermeneutic question that the nature of these commitments becomes apparent.

7

To see what is distinctive about an expressivist view on the meaning of moral language, I want to sketch an answer to the hermeneutic question that is congenial to expressivism's core commitment. It helps however to start by briefly looking at a different answer to the hermeneutic question, which relies heavily on representational notions. This *representationalist* answer is implicit in much theorizing about meaning—which explains why expressivism is often thought to conflict with standard semantic theory—but it is one that expressivist can and should want to reject.

The most straightforward version of the representationalist answer takes the notion of truth that figures in the semantic theory as being something like 'correspondence with reality'. Part of what it is for 'Bill was tortured' to mean

that Bill was tortured is for there to be some isomorphism between the structure of the sentence and the 'metaphysical structure' of some corresponding 'chunk' of reality.[38] This is not the place to question the inteligibility of such a view. At least prima facie, it is a way of reading off some substative claims about the relationship between linguistic items and features of the world so as to account for what an assignment of a set of worlds to that sentence is telling us about the world.

Another, less metaphysically loaded version of this answer takes the following form. What it is for the meaning of 'Bill was tortured' to be given by the set of worlds in which Bill was tortured is for speakers to utter that sentence in order to describe their environment, and for their description to be accurate just in case the concrete world the speaker lives in is an element of the corresponding set. Alternatively, we can say that what it is for the meaning of 'Bill was tortured' to be given by that set of worlds is for speakers to utter that sentence to indicate that they are in a state that represents the world they are in as being an element of that set—where 'representation' is understood so as to involve some kind of co-variation (see §1).

Whatever the merits of the representationalist answer to the hermeneutic question, it is not forced upon us.[39] This should not come as a surprise. Some think of mainstream semantics as assigning *propositions* to English sentences, where propositions are taken to be essentially representational. But semanticists are not in the business of pronouncing on metaphysical issues. It would be incredible if the viability of current semantic theory as we know it depended on the outcome of a controversial metaphysical dispute.

The representationalist answer to the hermeneutic question may sometimes appear to be part of the project of linguistic semantics as conceived by its practitioners. It might thus seem that a commitment to some form of methodological naturalism might settle the answer to the hermeneutic question. But deference on *this* question is no more warranted than deference to mathematicians on the question of mathematical platonism—or, for that matter, on the

[38] See the introduction to Price 2011 for an alternative, very vivid presentation of the view I have in mind.

[39] Cf. McDowell 1998, p. 484: "Sometimes [Sellars] suggests that the very idea of word–world relations as they figure in Tarskian semantics is 'Augustinian', in the sense that fits the opening sections of Wittgenstein's *Philosophical Investigations*. But this is simply wrong. It is perfectly congenial to Tarskian semantics to say that the notions of such word–world relations as denotation and satisfaction are intelligible only in terms of how they contribute to capturing the possibilities for 'making moves in the language-game' by uttering whole sentences in which the relevant words occur. These relations between words and elements in the extralinguistic order should not be conceived as independently available building blocks out of which we could construct an account of how language enables us to express thoughts at all." See also Davidson 1973.

hermeneutic question for probability. If a survey were to reveal that probability theorists generally espouse a subjectivist interpretation of probability would not settle the hermeneutic question for probability. Similarly, finding out that semanticists by and large endorse something like a correspondece theory of truth would not, by itself, settle the question of semantic hermeneutics.

As it happens, some semanticists explicitly disavow any commitment to the representationalist answer to the hermeneutic question.[40] But even if there weren't any, it is far from obvious that the success of truth-conditional semantics at what it sets out to accomplish depends on a particular answer to the hermeneutic question.[41]

In the introductory chapter to their textbook on semantics for generative grammar, Heim and Kratzer cite approvingly a well-known passage in (Davidson 1967, p. 311) on what a truth-conditional semantic theory aims to accomplish:

> The theory reveals nothing new about the conditions under which an individual sentence is true; it does not make those conditions any clearer than the sentence itself does. The work of the theory is in relating the known truth conditions of each sentence to those aspects ('words') of the sentence that recur in other sentences, and can be assigned identical roles in other sentences. Empirical power in such a theory depends on success in recovering the structure of a very complicated ability—the ability to speak and understand a language. (Heim and Kratzer 1998, p. 2),

On this view, a truth-conditional semantic theory simply takes for granted that we have an account of what it is for the primitive terms in the language to mean what they do, and builds from that an account of what it is for more complex expressions to mean what they do. Or, to put it in terms of knowledge, it takes for granted what it is to know the meaning of the primitive expressions in a language, and builds from that an account of what it is to know the meaning of the language. The starting point of such a theory will need to be fleshed out some way or another. But it is not a task for *compositional* semantics. Indeed, compositional semantic theory is designed to work largely independently of how those details get worked out.[42]

[40] Cf. Partee 1988, p. 118: "[I]t is the *structure* provided by the possible world theory that does the work, not the choice of particular possible worlds, if the latter makes any sense at all." See also Portner 2009, p. 116f.

[41] Cf. Pietroski 2003, p. 246: "But so far as I can tell, the issues that animate current research in semantics are orthogonal to the question of whether *truth* values are really the valuations of sentences."

[42] It is worth adding that, while orthodox semantics does not yet include an account of what it is for primitive terms to mean what they do, orthodox semantics does say much about the meaning of primitive terms. In particular, in its assignment of semantic values to certain lexical items, textbook semantics will predict the validity of certain patterns of inference (e.g. Kratzer's lexical

8

The representationalist answer to the hermeneutic question has been implicit in much of the literature on metaethical expressivism. It is no surprise, then, that the case for INCOMPATIBILISM has seemed so compeling. Suppose we take it for granted that an assignment like

(3) $[\![\textbf{Torture is wrong}]\!] = \{w : \text{Torture is wrong in } w\}$

amounts to the claim that the meaning of 'torture is wrong' is constituted by what it represents the world as being. Suppose that, in addition, we take it for granted that the meaning of a sentence is determined by the mental state that is expressed by it. Then we seem to have no choice but to think that the belief that torture is wrong is representational.[43] Once we endorse the representationalist hermeneutics, a semantic theory that yields an entry like (3) will seem incompatible with expressivism.

An expressivist who wants to endorse mainstream truth-conditional semantics had better provide an alternative to the representationalist hermeneutics, one that is compatible with her core commitments. Such an alternative answer could take the following form:

> Part of what it is for the meaning of 'Torture is wrong' to be given by the set of worlds in which Torture is wrong, is for that particular abstract object to adequately characterize certain relevant features (what I will call 'linguistically relevant features') of the mental state expressed (see §1) by an utterance of 'Torture is wrong',

Of course, this will only work if being a representational mental state is not a linguistically relevant feature of my belief that torture is wrong. For standard semantics does not mark a difference between 'Torture is wrong' and 'Running is tiring', and the expressivist will want to say that the mental state associated with the latter is representational, whereas the one associated with the latter is not.

So which are the 'linguistically relevant features' of the mental states in question? We cannot settle that question in advance of theorizing. But one reasonable guess is that the linguistically relevant features of a mental state will be those that play a role in explaining the communicative effect of an utter-

semantics for modals). Moreover, some stories about how best to develop a compositional semantic theory for English may have non-trivial consequences for doing lexical semantics (cf. Partee 1995).

[43] There may be ways of blocking this move—see e.g. Kalderon 2005b.

ance of that sentence.[44] The property of entailing that someone was tortured is a property of my belief that Bill was tortured which plays a role in explaining facts about linguistic behavior (e.g. that if I utter 'Bill was tortured' in a particular conversation, I would not follow that up with 'and someone was tortured'). The property of having been acquired on a Tuesday, say, presumably does not.[45]

Depending on our theory of conversational dynamics, different features of the state expressed by an utterance of 'Bill was tortured' will be relevant for the explanation of that utterance's conversational effects. But there is much we can agree on before settling on a particular such theory. It is in part because of what 'Bill was tortured' means that an utterance of that sentence has the effect that it has in a conversation—e.g. that it normally leads to changes in the attitudes of participants in a conversation in systematic ways; that utterances of other sentences ('Bill was not tortured') become unacceptable in the conversation unless the initial utterance is challenged; and so on.

We need to say what it is about the state expressed by an utterance of 'Bill was tortured' that accounts for these effects. And we need to say what it is about the state expressed by an utterance of 'Torture is wrong' that accounts for *that* utterance's conversational effects. But any difference among those states that does not play a role in explaining their contribution to a theory of conversational dynamics, on this view, will not need to be marked by our semantic theory. If the differences between the relevant states are visible only at a different theoretical level—say at the level of giving a fully general theory of the mind—the expressivist can maintain that there is an important difference among the mental states that occupy the roles of conversational states. Some such states may be representational, and some may not be. But unless

[44] Cf. Higginbotham 1992, p. 5: "The facts that semantics must account for comprise the context-independent features of the meaning of expressions that persons must know if they are to be competent speakers of the languages to which those features are assigned." I am assuming that, in order to be a competent speaker of English, one must at least be able to update one's attitudes during a conversation in particular ways. Being privy to the conversational dynamics for English is at least a necessary condition for being a competent speaker of English.

[45] Unlike the property of having been acquired on a Tuesday, the property of being non-representational is, if expressivism is true, an essential property of the belief that torture is wrong. As such, one might object, we should be able to read it off from the assignment of semantic value to 'torture is wrong'. But what motivates the idea that the semantics should be indifferent to the difference between states of mind that were assigned on a Tuesday and those that were not is that this is something that is not reflected in our use of language. The expressivist could grant that the state of mind expressed by 'torture is wrong' could have been a representational state—something like what non-naturalists think is the state of believing that torture is wrong. But if things had turned out that way, the dynamics of moral talk would have been much like what it actually is, assuming expressivism is true (although the dynamics of meta-ethical talk may well have been very different).

this distinction is relevant to the explanatory agenda of compositional semantics, whether a state of mind is representational will have no bearing on the compositional semantics.

The expressivist's hypothesis is that the theory of conversational dynamics will be the same *regardless* of whether conversational states are representational or not. If this is right, the differences between moral and non-moral thought will not be reflected at the level of the compositional semantics. The theory of conversational dynamics will specify certain roles for mental states expressed by utterances to play. And as long as one's full account of the relevant states implies that those states can serve as a realizer for that role, the theory of conversational dynamics will be compatible with that account of the relevant states. It is only once we look at features of the realizers that our theory of communication is blind to that the differences between the expressivist and her opponent will turn up.

I cannot offer a defense of this hypothesis here. But I can provide some reasons for thinking that something like it has got to be right.

Consider a picture of conversational dynamics along the lines of (Stalnaker 1978). On this picture, sentences are assigned sets of possible worlds so as to characterize their effects on the attitudes of participants in a conversation. The simplest way of implementing this is by using sets of possible worlds to characterize the attitudes of the participants, and to specify the effects of utterances on those states in terms of set-theoretic operations.

For example, if you are in the state of wondering whether Bill was tortured, we can characterize your state of mind with a set that contains worlds in which Bill was tortured and worlds in which he was not. If we characterize the state of believing that Bill was tortured using a set containing only worlds in which Bill was tortured, and we think that my uttering 'Bill was tortured' will get you to believe that Bill was tortured, we can assign to that sentence the set of worlds in which Bill was tortured. The effect of an utterance of that sentence can thus be characterized as the result of intersecting the set that characterizes the state you were in before my utterance with the set assigned by the semantics to the sentence uttered.

On the picture of conversational dynamics described above, the descriptivist is appealing to an algebra of sets of possible worlds. Each one of our moral and non-moral beliefs, on this picture, gets assigned an element of that algebra in such a way that set-theoretic relations correspond to inferential relations among beliefs in the usual way (set-theoretic inclusion corresponds to entailment, etc.). And each sentence of English is assigned, by the descriptivist's compositional semantics, an element of the algebra.

These possible worlds are typically identified as states of the world, in a way that seems unfriendly to expressivism. But this is not essential to the semantic

machinery. The 'possible worlds' in the semantics are abstract points. You can give a different understanding of what these points are without changing the structure of the semantics.

To see that, suppose that, for the purposes of modeling moral and non-moral thought, as well as the inferential relations among the relevant states, the expressivist gives us a theory on which: (i) each state of mind is associated with an element in an algebra A_E of sets; (ii) the set-theoretic relations among those elements in the algebra correspond to inferential relations among the states of mind in the usual way; and (iii) there is an isomorphism f from the resulting algebra to the algebra A_D used by the *descriptivist* to model our moral and non-moral thought. Now suppose the descriptivist has an adequate semantic theory which assigns, to each sentence s of English, an element $[\![s]\!]$ of A_D in such a way that s expresses the state of mind corresponding to $[\![s]\!]$. The expressivist can then simply adopt the descriptivist semantics, just by adding the following caveat: s expresses the state of mind corresponding to $f^{-1}([\![s]\!])$—that is, the state of mind corresponding to that element of A_E that gets mapped to the state corresponding to $[\![s]\!]$ by the isomorphism f.

Expressivists can understand those possible worlds in terms of what Gibbard (2003) calls 'fact-plan worlds'. A fact-plan world, on Gibbard's picture, can be thought of as a pair $\langle d, p\rangle$ consisting of a state of the world and a plan for action.[46] Gibbard uses sets of fact-plan worlds to characterize states of minds in something like the following way. Say that a fact-plan world is compatible with your view on what the world is like just in case, if you believe the world is such and such, the first element of the fact-plan world is one in which such and such. And it is compatible with your view on what to do just in case, for each ϕ and c you plan to ϕ in circumstances c, the second element of the fact-plan world has you ϕ-ing whenever you are in c. Your state of mind will now be characterized by the set of fact-plan worlds compatible with your view of the world and with your view on what to do.

The crucial point here is that the algebra of fact-plan worlds is isomorphic to the algebra of possible worlds.[47]

[46] For our purposes we can think of a plan as a function that determines, for any context, a particular course of action. This leaves out many of the subtleties in Gibbard's discussions of plan-laden thought, but it will do for now.

[47] Cf. Gibbard 2003, p. 47–58. Here is a proof of that claim, for the sake of completeness. Let W be the set of possible worlds, and assume the algebra A_D is just the collection of all subsets of W. Define two equivalence relations on possible worlds as follows: $w \sim_1 w'$ iff w and w' agree on all purely non-moral facts; $w \sim_2 w'$ iff for all contexts, w and w' agree on what action the agent of the context ought to do. (I'm assuming the descriptivist will want $\sim_1$ to be a non-trivial equivalence relation, even if she grants that moral facts supervene on non-moral facts. In some sense, then, some elements of W may have to be 'epistemically possible', if not metaphysically possible.) Denote by W_1 (resp. W_2) the set of equivalence classes under W_1 (resp. W_2). The

For the purpose of characterizing the motivational character of mental states, the finer structure of the sets of fact-plan worlds becomes important. For example, whether you can tell if a fact-plan world is in a set without knowing anything about its 'plan' component will determine whether the state characterized by that state is motivationally neutral. But the semantic machinery is not sensitive to the fine structure of these objects—in other words, it is not sensitive to the differences in the roles played by p and d in characterizing an agent's state of mind.[48]

9

Part of my goal has been to single out a particular kind of metasemantic question. That question—what I called the *hermeneutic* question—asks for an account of what it is for a particular semantic theory to be correct.

How to answer the hermeneutic question is not settled by which semantic theory happens to be correct. We should recognize that views can agree on what the *compositional semantics* of a fragment of language is, while disagreeing on how to *interpret* that compositional semantics. This is in keeping with a way of thinking of formal semantics as a modeling enterprise. As van Eijck and Visser 2010 put it (speaking of dynamic semantics in particular, but the point surely applies more generally): "[formal semantics] aims to *model* meaning and interpretation. You can do that without answering broader philosophical questions, such as the question what it is that makes it possible for the subject to be related to these meanings at all." As I would put it, we can model meaning without answering broader hermeneutic questions, such as the question of what it is for expressions in the language to have the meaning properties that they do.

Once we distinguish between the project of giving a formal semantics for English from that of answering the hermeneutic question for that theory, we make room for a different way of conceptualizing the expressivist's commitments in the theory of meaning. We can grant that expressivism has substan-

function that maps each w to $\langle [w]_1, [w]_2 \rangle$ (where $[w]_n$ is the equivalence class of w under $\sim_n$) is an isomorphism from the algebra of subsets of W to the algebra of subsets of $W^* = W_1 \times W_2$. And, clearly, the algebra of sets of W^* is isomorphic to the algebra of sets of fact-plan worlds.

[48] Things may turn out to be less straightforward when giving a semantics for attitude verbs. Yalcin 2012 gives one such semantics that makes apparent how, if we assign to moral and non-moral sentences semantic values of different kinds, we *can* nonetheless have a unique entry for 'believes' that can take prejacents of each type. But while it may well be possible to do so, the question I'm interested in is whether the expressivist needs to mark the differences between moral and non-moral discourse in the semantics. I have tried to make the case that there is none, but it may well be that, all things considered, we would gain some clarity by doing semantics in a slightly revisionary way.

tive commitments for a theory of *meaning*, while maintaining that it is (by and large) neutral on questions about compositional semantics.

This is not to say that expressivism's commitments in the theory of meaning are not costly. There are no doubt certain questions about explaining the structure of moral thought that become apparent once we see the shape that a semantic theory is going to take. A truth-conditional semantic theory, for example, will predict that certain sentences stand in logical relations to one another. It will predict, to use a well-known example (Dorr 2002), that the following is a valid argument:

P1. If lying is wrong, the souls of liars will be punished in the afterlife.
P2. Lying is wrong.
C. The souls of liars will be punished in the afterlife.

The expressivist owes us an account of the nature of moral thought that makes it apparent that our moral beliefs do stand in those relations to one another and to our non-moral beliefs. In particular, she owes us a story on which it can be rational to be in a position (e.g. that of accepting P1) to be disposed to change one's non-moral beliefs upon changing one's moral views.[49]

To be sure, that is a problem expressivists need to solve regardless whether or not INCOMPATIBILISM is true. And that is where the action should be. Expressivism may not, in the end, be the correct metaethical theory. But whether that is so will not be settled by questions in compositional semantics.[50]

REFERENCES

Almog, J. 1984. Semantical Anthropology. *Midwest Studies In Philosophy* 9.1, pp. 478–489.

Bennett, K. 2009. Composition, Colocation, and Metaontology. In: Chalmers, D., D. Manley, and R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press, pp. 38–76.

[49] While the problem of accounting for the rationality of inference is not new, I take Dorr's example to raise an even more pressing challenge, viz. that of accounting for rational transitions involving moral and non-moral beliefs. This, however, is not the place to argue that this is a distinctive challenge.

[50] Thanks to Matthew Chrisman, Gabriel Greenberg, Chris Meacham, Eliot Michaelson, Huw Price, Paolo Santorio, Mark Schroeder, and Michael Williams for illuminating comments and conversations. Thanks also to audiences at the Johns Hopkins University, UCLA, the University of Edinburgh, the University of Massachusetts at Amherst, and the University of Sydney. Special thanks to Alexi Burgess, Mark Greenberg, Brett Sherman, Katia Vavova, Kenny Walden, and Seth Yalcin for extremely helpful comments on earlier versions of this paper. Support from the Provost Postdoctoral Program at the University of Southern California, and from the Pragmatics Foundations Project, led by Huw Price at the University of Sydney, is gratefully acknowledged.

Björnsson, G. and A. Båve. 2007. Meaning as a Normative Concept. *Theoria* 73.3, pp. 190–206.
Blackburn, S. 1988. Attitudes and Contents. *Ethics* 98.3, pp. 501–517.
———. 1993. *Essays in Quasi-Realism*. Oxford: Oxford University Press.
———. 1998. *Ruling Passions*. Oxford University Press.
———. 2005. Quasi-Realism no Fictionalism. In: Kalderon, M. E. (ed.), *Fictionalism in Metaphysics*. Oxford: Oxford University Press, pp. 322–338.
Boghossian, P. A. 1990. The Status of Content. *Philosophical Review* 99.2, pp. 157–184.
Burgess, A. 2011. Mainstream semantics + deflationary truth. *Linguistics and Philosophy* 34.5, pp. 397–410.
Chomsky, N. 1995. Language and nature. *Mind* 104.413, pp. 1–61.
Chrisman, M. 2008. Expressivism, Inferentialism, and Saving the Debate. *Philosophy and Phenomenological Research* 77.2, pp. 334–358.
———. 2012. On the Meaning of 'Ought'. In: Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics*. Vol. 7. Oxford: Oxford University Press, pp. 304–332.
Davidson, D. 1967. Truth and meaning. *Synthese* 17.1, pp. 304–323. Reprinted in Davidson 1984, pp. 17—42.
———. 1973. In Defense of Convention T. In: Leblanc, H. (ed.), *Truth, Syntax, and Modality*. Vol. 68. Studies in Logic and the Foundations of Mathematics. Elsevier, pp. 76–86. Reprinted in Davidson 1984, pp. 65–76.
———. 1974. Belief and the basis of meaning. English. *Synthese* 27.3-4, pp. 309–323.
———. 1984. *Inquiries Into Truth And Interpretation*. Oxford University Press.
Davis, W. A. 2002. *Meaning, Expression, and Thought*. Cambridge: Cambridge University Press.
Dorr, C. 2002. Non-cognitivism and Wishful Thinking. *Noûs* 36.1, pp. 97–103.
Dreier, J. 2004. Meta-Ethics and The Problem of Creeping Minimalism. *Philosophical Perspectives* 18.1, pp. 23–44.
Dummett, M. 1975. What is a Theory of Meaning? (I). In: Guttenplan, S. (ed.), *Mind and Language*. Oxford: Oxford University Press. Reprinted in Dummett 1993, pp. 1–33. Page numbers refer to the reprinted version.
———. 1991a. *The Logical Basis of Metaphysics*. Cambridge, Mass.: Harvard University Press.
———. 1991b. "The Relative Priority of Thought and Language". In: *Frege and Other Philosophers*. Oxford University Press, pp. 315–324.
———. 1993. *The Seas of Language*. Oxford: Clarendon Press.
Eijck, J. van and A. Visser. 2010. Dynamic Semantics. In: Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2010.
Eklund, M. 2011. Fictionalism. In: Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2011.
Fine, K. 2012. Guide to Ground. In: Correia, F. and B. Schnieder (eds.), *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge University Press, pp. 37–80.
Fodor, J. A. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press.
Gallistel, C. R. and A. P. King. 2009. *Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience*. West Sussex: Wiley-Blackwell.

Gibbard, A. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, Mass.: Harvard University Press.
———. 2003. *Thinking How to Live*. Cambridge, Mass.: Harvard University Press.
Greenberg, M. 2005. A New Map of Theories of Mental Content: Constitutive Accounts and Normative Theories. *Philosophical Issues* 15.1, pp. 299–320.
Grice, H. P. 1957. Meaning. *Philosophical Review* 66.3, pp. 377–388.
———. 1969. Utterer's Meaning and Intention. *Philosophical Review* 78.2, pp. 147–177.
Hájek, A. 2012. Interpretations of Probability. In: Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. Summer 2012.
Harman, G. 1974. Meaning and Semantics. In: *Semantics and Philosophy*. New York: New York University Press, pp. 1–16. Reprinted in Harman 1999, pp. 192—205. Page numbers refer to the reprinted version.
———. 1999. *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press.
Heck, R. 2007. Use and Meaning. In: Auxier, R. E. and L. E. Hahn (eds.), *The Philosophy of Michael Dummett*. Chicago: Open Court, pp. 531–57.
Heim, I. and A. Kratzer. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.
Higginbotham, J. 1985. On Semantics. *Linguistic Inquiry* 16.4, pp. 547–593.
———. 1992. Truth and Understanding. *Philosophical Studies* 65.1/2, pp. 3–16.
Kalderon, M. E., (ed.). 2005a. *Fictionalism in Metaphysics*. Oxford: Oxford University Press.
———. 2005b. *Moral fictionalism*. Oxford: Oxford University Press.
Kaplan, D. 1989. Afterthoughts. In: Almog, J., J. Perry, and H. Wettstein (eds.), *Themes From Kaplan*. Oxford University Press, pp. 565–614.
Kraut, R. 1993. Robust deflationism. *Philosophical Review* 102.2, pp. 247–263.
Kripke, S. A. 1980. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.
Lewis, D. 1970a. General semantics. *Synthese* 22.1, pp. 18–67. Reprinted, with a postcript, in Lewis 1983, pp. 189–232. Page numbers refer to the reprinted version.
———. 1970b. How to Define Theoretical Terms. *Journal of Philosophy* 67.13, pp. 427–446. Reprinted in Lewis 1983, pp. 78—95.
———. 1975. Languages and Language. *Minnesota Studies in the Philosophy of Science* 7, pp. 3–35. Reprinted in Lewis 1983, pp. 163–189.
———. 1983. *Philosophical Papers*. Vol. 1. New York: Oxford University Press.
———. 1991. *Parts of Classes*. Oxford: Basil Blackwell.
———. 2005. Quasi-Realism is Fictionalism. In: Kalderon, M. E. (ed.), *Fictionalism in Metaphysics*. Oxford: Oxford University Press, pp. 314–321.
MacFarlane, J. 2010. Pragmatism and Inferentialism. In: Weiss, B. and J. Wanderer (eds.), *Reading Brandom: On Making It Explicit*. London: Routledge, pp. 81–95.
McDowell, J. 1998. Intentionality as a Relation. *Journal of Philosophy* 95.9, pp. 471–491.
O'Leary-Hawthorne, J. and H. Price. 1996. How to stand up for non-cognitivists. *Australasian Journal of Philosophy* 74.2, pp. 275–292.

Partee, B. H. 1988. Possible worlds in model-theoretic semantics: a linguistic perspective. In: Allén, S. (ed.), *Possible worlds in Humanities, Arts, and Sciences: Proceedings of Nobel Symposium 65*. Berlin: Walter de Gruyter, pp. 93–123.

———. 1995. Lexical Semantics and Compositionality. In: Gleitman, L. R. and M. Liberman (eds.), *An Invitation to Cognitive Science*. Second edition. Vol. 1. MIT Press, pp. 311–360.

Pietroski, P. M. 2003. The Character of Natural Language Semantics. In: Barber, A. (ed.), *Epistemology of Language*. Oxford: Oxford University Press, pp. 217–256.

Portner, P. 2009. *Modality*. New York: Oxford University Press.

Price, H. 2004. Immodesty Without Mirrors: Making Sense of Wittgenstein's Linguistic Pluralism. In: Kölbel, M. and B. Weiss (eds.), *Wittgenstein's Lasting Significance*. London: Routledge, pp. 179–206. Reprinted in Price 2011, pp. 200-227.

———. 2011. *Naturalism without Mirrors*. New York: Oxford University Press.

Quine, W. V. O. 1964. *Word and Object*. Cambridge, Mass.: The MIT Press.

Rayo, A. 2013. *The Construction of Logical Space*. Oxford: Oxford University Press.

Ridge, M. Forthcoming. *Impassioned belief*. Unpublished typescript, University of Edinburgh.

Rosen, G. 1998. Blackburn's *Essays in Quasi-Realism*. *Noûs* 32.3, pp. 386–405.

———. 2010. Metaphysical Dependence: Grounding and Reduction. In: Hale, B. and A. Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press.

Schiffer, S. 1982. Intention-based semantics. *Notre Dame Journal of Formal Logic* 23.2, pp. 119–156.

Schroeder, M. 2008a. *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.

———. 2008b. Expression for Expressivists. *Philosophy and Phenomenological Research* 76.1, pp. 86–116.

Sinclair, N. 2009. Recent Work in Expressivism. *Analysis* 69.1, pp. 136–147.

Speaks, J. 2011. Theories of Meaning. In: Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. Summer 2011.

Stalnaker, R. C. 1978. Assertion. In: Cole, P. (ed.), *Syntax and Semantics*. Vol. 9. New York: Academic Press, pp. 315–322. Reprinted in Stalnaker 1999.

———. 1984. *Inquiry*. Cambridge, Mass.: Bradford books, MIT Press.

———. 1997. Reference and Necessity. In: Hale, B. and C. Wright (eds.), *A Companion to the Philosophy of Language*. Oxford: Blackwell, pp. 534–554.

———. 1999. *Context and Content*. Oxford: Oxford University Press.

———. 2002. Common Ground. *Linguistics and Philosophy* 25.5, pp. 701–721.

Stich, S. 1992. What Is a Theory of Mental Representation? *Mind* 101.402, pp. 243–261.

Wedgwood, R. 2007. *The Nature of Normativity*. Oxford: Oxford University Press.

Wiggins, D. 1972. On Sentence-Sense, Word-Sense, and Differences of Word-Sense. In: Steinberg, D. D. and L. A. Jokobovits (eds.), *Semantics*. Cambridge: Cambridge University Pres, pp. 14–34.

Williams, M. 1999. Meaning and Deflationary Truth. *Journal of Philosophy* 96.11, pp. 545–564.

Williams, M. 2010. Pragmatism, Minimalism, Expressivism. *International Journal of Philosophical Studies* 18.3, pp. 317–330.
Yalcin, S. 2007. Epistemic Modals. *Mind* 116.464, pp. 983–1026.
———. 2012. Bayesian Expressivism. *Proceedings of the Aristotelian Society* 112.2, pp. 123–160.