Negation, expressivism, and intentionality

Alejandro Pérez Carballo

University of Massachusetts, Amherst

Many think that expressivists have a problem with negation.[1] I disagree. Or at least, I disagree that *expressivists* have a problem with negation. For *if* there is a problem with negation, it is a problem most of us have—those of us, expressivists and descriptivists alike, who accept some plausible claims about the nature of intentionality.

Whether there is any special problem for expressivists turns, I will argue, on whether facts about what truth-conditions beliefs have can explain facts about basic inferential relations among those beliefs. And I will suggest that the answer to this last question is, on most plausible attempts at solving the problem of intentionality, *no*.

1   Introduction

According to metaethical expressivists, when I tell you that cannibalism is wrong I do not take a stance on what the world is like—I am not in the business of describing the world. In the jargon, expressivists deny that moral thought and talk is *representational.* And this, according to the critics, precludes expressivists from giving an explanation of the following fact:

> NEGATION: It is inconsistent to accept a moral sentence and its negation.

Some have taken this observation to raise a devastating problem for expressivists (see Schroeder 2008c, for discussion). The claim is that descriptivists can give an explanation of NEGATION in terms of the representational properties of the relevant sentences, and that this explanation is much better than any one that expressivists have to offer.

Against this, I will argue that if NEGATION is something that needs to be explained, then any adequate explanation available to the descriptivist will be available to the expressivist. So if expressivists have a problem with negation—if

---

1 E.g. Dreier 2006b,a, 2009, van Roojen 1996, Schroeder 2008c,a.

they have no adequate explanation of NEGATION—then it is a problem we all have.

Throughout I will lean heavily on a nontrivial assumption, viz. that there is a solution to the problem of intentionality. More specifically, I will assume that we can explain, in non-intentional terms, why words and thoughts have the representational properties that they do. I will not argue for that here—my conclusions should be read as conditional on this assumption. But even the conditional conclusions should be of interest, for this assumption is one that many descriptivists would be happy to endorse.

2 The problem

At the relevant level of abstraction, *expressivism* about moral discourse (e.g. Blackburn 1998, Gibbard 1990, Stevenson 1937) is a conjunction of two claims. First, that the meaning of declarative 'moral sentences' is a function of the mental states they are used to express. Second, that the relevant states of mind— and moral thought more generally—are non-representational. According to expressivism, to make a moral judgment does not involve taking a stance on what the world is like—rather, it involves being in a desire-like state of some kind.[2]

An example might help. When I claim that Tom is a cannibal I am describing a way the world is, viz. one in which Tom is a cannibal. The point of making this claim is to provide you with information about the world—to tell you that the world is this way. And when I judge that Tom is a cannibal, I am taking a stance on what the world is like—I am ruling out a way the world could be, viz. one in which Tom is not a cannibal.

According to *metaethical descriptivists*, the same is true of my claim that cannibalism is wrong. When I tell you that cannibalism is wrong I am describing a way the world is. I am conveying information—about a realm of irreducibly moral, *sui generis* facts, perhaps; or about some yet to be specified natural facts (involving our conventions, our desires, or what have you). And when I think that cannibalism is wrong, I am taking a stance on what the world is like—I am ruling out a way the world could be, viz. one in which cannibalism is not wrong.

In contrast, expressivists hold that when I tell you that cannibalism is wrong I am *not* describing a way the world is. The meaning of my utterance is thus not given by how it represents the world as being. Rather, its meaning is given by the fact that I thereby express some distinctive state of mind. And this state of mind—the state of thinking that cannibalism is wrong—is not a representational

---

2 It may be best to reserve the term 'expressivism' for the first of these two claims, and to use *non-cognitivism* as a label for second one. I am following standard usage in using 'expressivism' for the conjunction of these two claims, even though I think this obscures some important issues. Fortunately, those issues are irrelevant for our present purposes.

state.[3] Rather, to a first and rough approximation, it is the state of *disapproving* of cannibalism. The point of uttering 'cannibalism is wrong' is to evince my disapproval, and perhaps to get you to disapprove, of cannibalism.[4]

Now focus on the state of mind of accepting (1) and the state of mind of accepting (2):

(1)   Cannibalism is wrong.

(2)   Cannibalism is not wrong.

There is something defective about someone who is in both of these states—there is some kind of conflict, something like 'intrapersonal disagreement', among those attitudes. As I will put it, all parties agree that it is *inconsistent* to accept (1) while accepting (2).[5] What is at issue is whether expressivists can provide an explanation of this fact—more generally, whether expressivists can provide an explanation of NEGATION.

2.1   Candidate explanations

It could turn out that there is an explanation of NEGATION in purely normative terms. If so, that explanation will presumably be available to the expressivist. After all, hers is a view about the nature of moral thought and talk, one that should in principle be compatible with any first-order moral theory. According to the descriptivist, however, in order to explain NEGATION we need to appeal to some non-normative claims. In particular, we need to appeal to the representational properties of moral judgments.[6] And the problem is that, according to the expressivist, these are properties that moral judgments do not have.

To see this more clearly, let me go over what the descriptivist's explanation of NEGATION is supposed to be:[7]

---

3 At least not one which represents that cannibalism is wrong.

4 This distinction can be drawn in domains other than morality: e.g. epistemic modality, epistemic normativity, causality, probability judgments, mathematics, and, in a way, negation itself, to name a few—see, e.g. Yalcin 2011, Field 2009, Blackburn 1990, Price 1983, Pérez Carballo 2014, and Price 1990, respectively. Most of the issues raised here do not depend on the specific flavor of the expressivist/descriptivist distinction under consideration. But because of its familiarity, and because it is typically taken to be the *locus* of the negation problem, I will focus mainly on expressivism about moral discourse.

5 It is not obvious how to characterize this type of conflict in a way that is neutral between expressivism and descriptivism (cf. §3.1), but I will assume there is such a way. I will say that the attitudes are *inconsistent* with one another whenever they conflict with one another in this particular way.

6 Here I am taking for granted that facts about the representational properties of mental states are not normative. This is a substantial assumption, but one that plausibly makes things harder for the expressivist. I trust that I am not begging any questions here, but I realize that the issues are subtle—see e.g. Rosen 1997, Gibbard 1994, 2012 for discussion.

7 I will be assuming throughout that both parties agree on what Schroeder 2008b calls *mentalism*: the view that mental content is explanatorily prior to linguistic content. As Schroeder points out, mentalism is a non-trivial commitment of expressivism, one that descriptivists might deny. That

DESCRIPTIVIST EXPLANATION: To accept a moral sentence is to believe its representational content. The representational content of a declarative moral sentence is incompatible with that of its negation. But it is inconsistent to have beliefs with incompatible representational contents. So it is inconsistent to accept a moral sentence and its negation.

Accepting a moral sentence, on the descriptivist picture, is a matter of having a belief—of being in a state with a particular representational content. So the explanation the descriptivist offers presupposes that moral judgments are representational. Given her views on the nature of moral judgment, the expressivist cannot appeal to DESCRIPTIVIST EXPLANATION in order to explain NEGATION.

By itself, this observation does not rule out the possibility of a suitable explanation of NEGATION that is compatible with expressivism. But given what we have said so far about expressivism, it is far from clear what such an explanation could look like.

This is most clearly seen through an example. Recall that, according to our toy version of expressivism, the attitude expressed by (1) is that of disapproval of cannibalism, and the attitude expressed by (2) is that of tolerance of cannibalism. So in order to explain why it is inconsistent to accept (1) and (2), the expressivist needs to appeal to something like

DISAPPROVAL FORBIDS TOLERANCE: It is inconsistent to disapprove of and tolerate the same thing.

Thus, she needs to assume that there is inconsistency in holding *different* types of attitudes toward the *same* object.[8] The descriptivist, in contrast, need only assume that for some attitudes, there is inconsistency in holding *that* attitude toward incompatible contents:

BELIEF COHERENCE: It is inconsistent to have beliefs with incompatible representational contents.

And the difference here, the critics insist, is subtle but important. For it means that while BELIEF COHERENCE can be taken for granted in explaining NEGATION, DISAPPROVAL FORBIDS TOLERANCE cannot.[9] As Schroeder (2008b) puts it:

---

said, mentalism seems to be the majority view among philosophers of mind and language—see e.g. Grice 1957, 1969, Lewis 1975, Schiffer 1982, Stalnaker 1984, Davis 2002; notable dissenters include Davidson 1974 and Dummett 1991—and can be motivated independently of the debate between expressivists and descriptivists.

8 This is no mere artifact of the analysis of accepting (2) in terms of tolerance. There simply is no way of analyzing accepting (2) in terms of disapproval of anything that is incompatible with cannibalism. Cf. Unwin 1999.

9 Cf. Schroeder 2008c on the difference between '*A*-type' and '*B*-type' inconsistency.

All [the] good models of inconsistency between mental states arose in [...] cases of the same attitude toward inconsistent contents. But tolerance of murder and disapproval of murder are two distinct and apparently logically unrelated attitudes toward the same content. [...] Assuming that disapproval and tolerance of murder are inconsistent is taking for granted everything that expressivists need to explain. (p. 48)

Now, there is room for the expressivist to push back. For example, she could point out that in assuming that it is inconsistent to bear the belief relation to incompatible propositions, the descriptivist is taking for granted everything that needs to be explained. After all, there is no inconsistency in thinking *p* is somewhat likely and at the same time thinking not-*p* is somewhat likely. By relying on BELIEF COHERENCE, the descriptivist simply restated the problem: for now we can ask why the incompatibility of two propositions—or the relation between truth and falsity—should somehow give rise to a conflict among the relevant attitudes.[10]

Alternatively, she could point out that it seems inconsistent to have intransitive preferences: to prefer *A* to *B*, *B* to *C*, and *C* to *A*. But there is no sense in which the contents of these attitudes are incompatible. The kind of conflict among attitudes exhibited by disapproving and tolerating the same thing may well be of a kind with that exhibited by having intransitive preferences. And perhaps the conflict exhibited by having intransitive preferences is of a kind with that exhibited by having beliefs with incompatible contents. So it may be that DISAPPROVAL FORBIDS TOLERANCE has as good a claim as BELIEF COHERENCE to be taken for granted in an explanation of NEGATION.[11]

---

10 Cf. Price 1990. I do not mean to suggest that this line of response suffices to get the expressivist out of trouble. Surely the descriptivist could in turn point out ways in which *thinking p is somewhat likely* differs from *believing p*—the latter, but not the former, 'aims at the truth', for example—that would explain why beliefs, but not thoughts about what's somewhat likely, with incompatible contents are in conflict. But it would be a mistake to think *this* would clearly suffice to get the *descriptivist* out of trouble. For the notion of aiming at the truth may well turn out to be a normative one (although see Rosen 2001 for discussion, as well as the more recent overview of some of the relevant literature in McHugh & Whiting 2014). And once we allow for normative notions to carry the explanatory burden, it is no longer obvious that the resulting explanation is not available to the expressivist. Be that as it may, I will not be presupposing in what follows that there is any merit to this line of response on behalf of expressivists. (Thanks here to an anonymous reviewer.)

11 For a recent and ambitious attempt at relying on this type of conflict between attitudes to give an account of semantic and logical inconsistency, see Baker & Woods (2015). As Baker & Woods (2015, 423) point out, citing Restall (2005), a version of inferentialism in the philosophy of logic (known as 'bilateralism' or 'rejectivism') explains the meaning of logical constants in terms of both acceptance and rejection, in an attempt at reconciling classical logic with the plausible idea that the meaning of logical constants is determined by their introduction and elimination rules (see e.g. Smiley 1996, Rumfitt 2000). And this entire research program proceeds by taking something like DISAPPROVAL FORBIDS TOLERANCE—a type of conflict between different attitudes towards the same content—for granted in explaining the meaning of negation. See Incurvati & Schlöder 2019 for a fascinating proposal, one that builds on this 'bilateral' tradition, for how to account for logical inconsistency in expressivist-friendly terms.

But I do not intend to pursue this any further. Let us just grant the critic that it is a desideratum for an explanation of NEGATION that it only posit inconsistency among attitudes of the same type, and only when their contents are inconsistent. Can the expressivist provide an explanation meeting this desideratum?

On the face of it, it seems she can. True, our toy version of expressivism posits inconsistency among attitudes of different types.[12] But not all versions of expressivism need to do that. In particular, the account of normative thought developed by Gibbard (1990) has all the resources to formulate such an explanation, one that only posits inconsistency among attitudes of the same type.[13]

Start with a few definitions.[14] For our purposes, a (complete) *norm* is a partition of all possible courses of actions, relative to a given context, into three categories: those that are *required*, those that are *forbidden*, and those that are *merely permitted*—i.e., permitted but not required. A norm is thus essentially a function that assigns to each course of action, relative to a context, one of three possible values. We can now define a *system of norms* as simply a set of norms. Finally, say that an action is required (resp. forbidden or merely permitted) by a system of norms just in case every complete norms in the set counts the action as required (resp. forbidden or merely permitted).

On Gibbard's picture, (1) expresses acceptance of a system of norms that forbids cannibalism. And (2) expresses acceptance of a system of norms that permits cannibalism. Since every norm that does not forbid cannibalism will permit it, the system of norms corresponding to (1) will be disjoint from the one

---

12 Cf. fn. 8

13 Another view which is designed to have those resources is that of Schroeder 2008c,a. Schroeder introduces the attitude of *being for* and suggests that we take (1) and (2) as both expressing the attitude of being for, with (1) corresponding to *being for disapproving of cannibalism* and (2) corresponding to *being for not disapproving of murder*. Unfortunately, introducing this much internal structure into the relevant attitudes comes at a rather high cost. I do not have the space to rehearse the arguments for that conclusion. For that, see Schroeder 2008a, esp. ch. 12. In my view, some of the difficulties that Schroeder raises for expressivists rely on an assumption that I think expressivists should reject, viz. that 'the meaning of a sentence *is* the mental state that it expresses' (p. 177, my emphasis). But I cannot argue for that here.

14 A word of warning. Much of what I go on to say may well be incompatible with the expressivist picture in Gibbard 2003. In particular, whereas in Gibbard 1990 the objects of attitudes are sets of norms, each of which determines a three-way partition of courses of actions (see e.g. Gibbard 1990, p. 95), in Gibbard 2003 they are sets of *hyperplans*, each of which determines a *two* way partition of courses of actions. I will also steer clear of Gibbard's frequent gloss of a world-norm pair as 'fully opinionated credal-normative states' for I think it can be misleading. A norm, as I will understand it, is simply an abstract object that partitions possible courses of actions into three categories. What is needed for an adequate theory of normative thought is not a theory of what norms 'really' are, but a theory of what accepting a norm (or set thereof) amounts to. (Similarly, one can give an account of belief without taking a stance on what possible worlds 'really' are, as long as one gives an adequate account of what having a belief with some set of possible worlds as its content ultimately amounts to.)

corresponding to (2)—no complete norm can be in both a set of norms all of which forbid cannibalism and a set of norms all of which permit cannibalism.[15]

This can be generalized further, to give an account of what accepting any moral sentence amounts to. Rather than going over the details here, I will just skip ahead and stipulate that (i) for any moral sentence there is a well-defined system of norms corresponding to that sentence and that (ii) the system of norms assigned to a moral sentence is disjoint from that assigned to its negation.[16]

Using these resources, the expressivist can now give an explanation of NEGA-TION that is much like the one given by the descriptivist. Say that two systems of norms are incompatible just in case they have an empty intersection. To give an explanation of NEGATION of the same form as DESCRIPTIVIST EXPLA-NATION, we can replace DISAPPROVAL FORBIDS TOLERANCE with

> NORMATIVE COHERENCE: It is inconsistent to accept incompatible systems of norms.

Once we do that, we get something that looks very much like DESCRIPTIVIST EXPLANATION:

> EXPRESSIVIST EXPLANATION: To accept a moral sentence is to accept a system of norms corresponding to that sentence. The system of norms corresponding to a sentence is incompatible with the system of norms corresponding to its negation. But it is inconsistent to accept

---

15 This is, of course, an oversimplification. Gibbard's hypothesis is that the attitude of accepting a system of norms must ultimately be part of an adequate theory of human psychology. Gibbard spends the first two parts of the book sketching a hypothesis about what acceptance of a system of norms could be, and how it could fit within a plausible theory of human psychology. Doing full justice to his account would take us far beyond the scope of this paper.

16 The story is a bit more complicated, for two reasons. First, because every sentence in the language gets assigned a set of *pairs* consisting of a possible world and a norm. The interested reader should consult Chapter 5 of Gibbard 1990, for the basic idea. Second, because as Dreier 1999 points out (p. 571), the story Gibbard tells needs to be revised if we are to get the right assignment for sentences with normative predicates embedded in intensional contexts. (This is also the source of an objection to Gibbard's view in Unwin 2001—see pp. 62-65.) The problem, in a nutshell, is that on Gibbard's semantics, attitude verbs come out as quantifying over possible worlds. When embedding a clause whose semantic value is a set of pairs, the resulting sentence comes out as a non-purely descriptive claim. Intuitively, however, 'Tom thinks that cannibalism is wrong' is purely descriptive, even if the prejacent is not. Fortunately, the needed revisions are reasonably straightforward. All that is required is to interpret attitude verbs as quantifiers over world-norm pairs, much like in semantics for attitude reports in systems that replace sets of possible worlds with sets of centered worlds (for an overview, see Ninan 2010). This will ensure that 'Tom thinks that cannibalism is wrong' is true at a world-norm pair $\langle w, n \rangle$ iff 'Cannibalism is wrong' is true in every world-norm pair compatible with what Tom accepts in w—instead of what Gibbard's original recipe yields, which is that 'Tom thinks that cannibalism is wrong' is true at a world-norm pair $\langle w, n \rangle$ iff for every w' compatible with what Tom thinks, 'Cannibalism is wrong' is true at $\langle w', n \rangle$ (note that the same 'n' is on both sides of the biconditional).

incompatible systems of norms. So it is inconsistent to accept a moral sentence and its negation.

The point of the rest of the paper is to argue that, to the extent that DESCRIP-TIVIST EXPLANATION is a good explanation of NEGATION, so is EXPRES-SIVIST EXPLANATION.

2.2 Qualms about the expressivist's explanation

For those familiar with Gibbard's formalism it should not come as a surprise that expressivists can give an explanation of NEGATION with much the same *structure* as the one given by descriptivists. The problem many have with EX-PRESSIVIST EXPLANATION, as I understand it, has to do with the status of NORMATIVE COHERENCE. The concern is that in relying on NORMATIVE COHERENCE, expressivists are assuming the very thing that they were supposed to explain.

This sort of concern was voiced explicitly by Jamie Dreier, in a slightly different context (he is talking of interpersonal disagreement, rather than inconsistency):

The problem is that if we cannot say anything by way of explanation [of why accepting a system of norms is inconsistent with accepting its complement], then we are hostage to the possibility that the intuitive notion of disagreement that we rely on, that we are taking for our purpose [*sic*] as primitive, is not friendly to [expressivist theories]. Maybe the intuitive notion is this: when we can see, by our native grasp of our language, that your rejection and my acceptance of this certain sentence counts as disagreement, that is because we have a prior grasp on the idea that some sentence really express real propositions [...] (Dreier 2009, p. 107)

Transposing this to the current context, the thought is this. In presupposing that there are attitudes (accepting a system of norms) that relate to one another much like beliefs do—attitudes that bear logical or inferential relations to one another that mimic those that hold among beliefs—the expressivist may be relying on assumptions she is not entitled to.

The worry is that the notion of *inconsistency*—and logical or inferential notions more generally—is best explained, in the case of beliefs, in terms of beliefs' representational properties. And unless the expressivist has an alternative explanation of how moral judgments can stand in logical relations to one another, the truth of claims like NORMATIVE COHERENCE may turn out to be evidence against the very core claims the expressivist wants to make. As Paul Horwich puts it—summarizing (Smith 1994)—the consensus is that 'the expressivist owes us some explanation of how it is possible for evaluations to function inferentially *as if* they were beliefs' (Horwich 1994, p. 19).

Strictly speaking, this is not the only sort of complaint that has been leveled against the sort of expressivist-friendly explanation I sketched above. Famously, Nicholas Unwin has argued that expressivists are unable to distinguish what it is to think that cannibalism is not wrong from what it is *not* to think that cannibalism is wrong.[17] Unwin's concern, at least as it applies to the view formulated in Gibbard 1990, arises quite naturally in light of Gibbard's reliance on the somewhat opaque notion of 'ruling out'. As Gibbard would put it, to think that cannibalism is not wrong just is to 'rule out' thinking that cannibalism is wrong. But, the worry goes, a committed agnostic rules out thinking that cannibalism is wrong without thereby thinking that cannibalism is *not* wrong—for she also rules out thinking that cannibalism is not wrong. So we are left without the ability to distinguish the agnostic from someone who thinks that cannibalism is wrong. Unwin considers a number of different responses on Gibbard's behalf, but finds them all wanting. But one he does not consider (later offered in Gibbard 2003, pp. 72ff) is that the relevant notion of 'ruling out' should be understood in terms of disagreement (or inconsistency). If that response is available to the expressivist, she can maintain that the difference between thinking that cannibalism is not wrong and not thinking that cannibalism is wrong is that in the former, but not the latter, one disagrees with thinking that cannibalism is wrong.

On the expressivist view sketched above, someone who thinks that cannibalism is not wrong is someone whose state is characterized by a set of norms none of which forbid cannibalism. And someone who does not think that cannibalism is wrong is someone whose state is characterized by a set of norms *not all* of which forbid cannibalism. The formalism clearly draws a distinction between the two states. And to see that this is not a distinction without a difference, it suffices to point out that the former state, but not the latter, is *inconsistent* with thinking that cannibalism is wrong. To be sure, one might find have methodological qualms about appealing to the notion of disagreement to characterize the distinction between what thinking that cannibalism is not wrong and not thinking that cannibalism is wrong. But this worry is ultimately one about the status of NORMATIVE COHERENCE.[18]

---

17 See Unwin 1999, 2001. See also Dreier 2006b, who worries that the expressivist cannot tell us a story as to the difference between 'indecision' and 'indifference' about whether cannibalism is wrong—the difference between someone who accepts *neither* that cannibalism is wrong nor that cannibalism is not wrong and someone who thinks cannibalism is merely permitted.

18 All of that said, this is not the only thing the expressivist can say. Following Gibbard 1990, we can say that to accept a system of norms is to be in a certain dispositional state. Simplifying a fair bit (in particular, eliding the distinction between internalizing and accepting a norm), we could say that to accept a system of norms all of which forbid cannibalism is to be disposed to both feel guilt upon engaging in cannibalism and to feel resentment towards others who engage in cannibalism. And to accept a system of norms none of which forbid cannibalism is to be disposed both to not feel guilt upon engaging in cannibalism and to not feel resentment towards others who engage in cannibalism. Someone who is agnostic is just one who lacks both dispositions. (Compare this to

3 Consistency and the attribution of content

All parties in the debate agree that it can be inconsistent to have a certain pattern of beliefs. But according to the expressivist, beliefs are not alone in that. For on her view there are states of mind that lack representational content which can be inconsistent with one another as well.

This should be cause for alarm only if something like the following claim is true:

> PRIORITY OF CONTENT: Beliefs are inconsistent with one another in virtue of the way their representational contents relate to one another.

For then the question would arise as to what explains why states that don't have representational content could be inconsistent with one another much like beliefs can be. If PRIORITY OF CONTENT is true, then BELIEF COHERENCE might seem like a safe thing to appeal to in explaining NEGATION, whereas NORMATIVE COHERENCE might seem to cry out for an explanation.[19] In short, if PRIORITY OF CONTENT is true, then expressivists face an explanatory challenge that descriptivists do not. But do we have reason to think that PRIORITY OF CONTENT is true?

My goal in this section is to argue that we do not. Before doing that, it is worth taking a moment to understand what PRIORITY OF CONTENT amounts to.

3.1 Inconsistency and irrationality

There are two notions in PRIORITY OF CONTENT in need of clarification: the 'in virtue of' locution, and the notion of inconsistency.

According to PRIORITY OF CONTENT, facts about what contents states have are *prior* to facts about which states are inconsistent. Priority here should be understood as *explanatory* priority: it is (partly) because two beliefs have contents that relate to one another in a particular way that the beliefs come to be inconsistent with one another. This, I submit, is the best way of interpreting PRIORITY OF CONTENT so as to raise a *prima facie* problem for expressivists.

I expect mention of relations of metaphysical dependence will make some nervous. So let me just reiterate that in order to make sense of the negation problem, we need to make sense of the notion of 'in virtue of', or the related notion of metaphysical explanation. The problem, recall, is based on an explanatory

---

what a dispositional account of belief would say about the difference between someone who has an opinion on whether it is raining and someone who is agnostic about it.)

19 Cf. Schroeder 2008c, p. 585: "By assigning the complement set of [complete systems of norms] to a negated sentence, all that Gibbard's account does, is to stipulate that it is to express a state of mind that is inconsistent with the state of mind expressed by the original sentence. But it does nothing to tell us […] why it is inconsistent with the state of mind expressed by the original sentence."

challenge. The challenge is that representational contents can play an *explanatory* role that systems of norms cannot. And the notion of explanation here cannot be causal. To the extent that the relevant notions of priority are unintelligible, so is the negation problem. Here, I will simply grant the critic the intelligibility of these notions. In so doing, I am only making the expressivist's task more difficult.[20]

As for the notion of inconsistency, let me just say something about what it cannot be, if PRIORITY OF CONTENT is to have a chance of being true.

Inconsistency cannot be defined so that two states are inconsistent just in case they cannot be true at the same time. This is for two reasons. First, because the challenge for the expressivist is that of explaining something she herself takes to be true. But if only states with representational content can be inconsistent, then the expressivist would just deny that the states of accepting (1) and (2) are inconsistent.[21] Second, and most importantly, because if that is how we *define* inconsistency, then PRIORITY OF CONTENT will be trivially false.[22]

To see that, note that according to PRIORITY OF CONTENT we can *explain* why two beliefs are inconsistent by saying that their representational contents are incompatible—i.e. that their contents cannot be true. But this is just another way of saying that we can explain why two beliefs are inconsistent by saying that they cannot both be true at the same time. And if as a matter of definition two beliefs are inconsistent just in case they cannot both be true at the same time, PRIORITY OF CONTENT would boil down to the trivially false claim that we can explain why two beliefs cannot be true at the same time by saying that they cannot both be true at the same time.

How else should we understand inconsistency? One option is to tie inconsistency very closely with irrationality. Here's how such a proposal could go. If two beliefs are inconsistent, then it is *irrational* to be in both states at once. So a necessary condition for two beliefs to be inconsistent is that it be irrational to have both beliefs at once. Now, this cannot be a sufficient condition for inconsistency. It may be irrational to be in a so-called Moore-paradoxical state: it

---

20 For a defense of the intelligibility of these notions, see e.g. Rosen 2010.

21 I have done my best to avoid the debate over whether expressivist can say that our moral talk and thought can be true—or have representational content—in some 'minimal' sense. In part, this is because I suspect that if we understand inconsistency in terms of incompatible truth-conditions, and we embrace some form of minimalist conception of truth, then NEGATION is not something we need to explain. But also, this is because there are some subtle issues here—e.g. whether expressivism is compatible with minimalism—that are beyond the scope of this paper. On this point, see e.g. Smith 1994, Dreier 2004.

22 What if we replaced PRIORITY OF CONTENT with something like: 'For two beliefs to be inconsistent just is for their representational contents to be incompatible'? The problem for the expressivist, on this view, would be that of giving a definition of inconsistency that draws the line where we expect it to be—after all, she cannot use the definition in terms of representational content. But expressivists can simply give a definition of inconsistency in terms of incompatibility of system of norms: two states of mind are inconsistent just in case the systems of norms that characterize them have an empty intersection.

may be irrational for me to believe that it is raining while believing that I do not believe that it is raining. But intuitively there is no inconsistency in having both of those beliefs at once.[23] Still, we could say that inconsistency is a particular *kind* of irrationality.[24]

One problem with this proposal is that it would rule out by definition views on which it is sometimes rational to have inconsistent beliefs. We might think that a lesson of the Paradox of the Preface is that it is sometimes rational to have inconsistent beliefs. But even if we are willing to accept that inconsistency always involves irrationality, the main problem with this proposal is that, pending further work, it is not very illuminating. For it fails to characterize what is distinctive about inconsistency as opposed to any other type of irrationality.

Perhaps the thing to say is that inconsistency is a primitive notion. Plausibly, whatever inconsistency amounts to, it had better be one of those relations among mental states that are the object of our theories of rationality—say, the kind of relation among mental states that is relevant for a computational or inferential characterization of an agent's cognitive economy. Beyond that, however, there simply is not much more to say about what it is for two mental states to be inconsistent. This still leaves it open whether there is an explanation of what makes it the case that two states are inconsistent with one another. But even if there is, such an explanation would not give an analysis of inconsistency.

What bearing does this have on the status of PRIORITY OF CONTENT? I have assumed that PRIORITY OF CONTENT is a substantive claim. One lesson from the preceding is that for this to be so, we need an independent grip on the notion of inconsistency—either as a primitive notion, or as some restricted type of irrationality. This means that it is in principle possible that this notion of inconsistency will not be explained in terms of the notion of representational content, and that instead the notion of inconsistency will play a role in the explanation of why beliefs have the representational content that they have.

3.2   Against Priority of Content, I

What makes it the case that a particular thought of mine has the representational content that it does? Call this the *problem of explaining content*—otherwise known as *the problem of intentionality*. Presumably we all want to solve this problem—descriptivists and expressivists alike.[25] If it turned out that PRIORITY

---

23 For a more detailed argument against a straightforward analysis of logical relations among mental states in terms of the rationality or irrationality of being in those states at once, see van Roojen 1996.

24 This line of thought is very much in line with the Fregean idea that the '[l]aws of logic […] are the most general laws, which prescribe universally the way in which one ought to think if one is to think at all'. Frege 1893, p. 12.

25 Of course, descriptivists need a solution to the problem that carries over to the case of moral beliefs, whereas expressivists would be content with a solution to the problem that explains the content of our non-moral beliefs. (Indeed, I believe one of the best motivations for metaethical expressivism is

OF CONTENT gets in the way of doing so, we would have a very good reason to reject it.

What could this explanation of content look like? Here is a possible candidate. Beliefs and desires are correlative dispositions to act.[26] To believe that Tom is a cannibal is to be disposed to act in a way that would satisfy one's desires in a world in which Tom is a cannibal and all of one's other beliefs are true. Similarly, to desire that cannibalism be outlawed is to be disposed to act in ways that would tend to bring it about that cannibalism is outlawed in worlds in which one's beliefs are true.[27]

The proposition that gets assigned as the content of your belief that Tom is a cannibal is an abstract object that is used to characterize your behavioral dispositions (ditto for your desire that cannibalism be outlawed). I relate to the proposition that Tom is a cannibal in virtue of having the belief that Tom is a cannibal. But this is not some mysterious, non-natural relation, involving some special faculty that gives me access to the realm of abstract objects. Rather, it is much like the way in which a chair can relate to the number 4.2 in virtue of weighing 4.2 pounds.[28] There is no mystery as to how a chair can bear the *weight in pounds* relation to an abstract real number—there is no need to assume that my chair has some super-natural faculty that allows it to reach out to some platonic realm.

This way of thinking about content attribution—where propositions are used to characterize our mental states much like numbers are used to characterize weights—has much to recommend it. But for our purposes it is enough to point out this much: something like this may be our best shot at solving the problem of intentionality, viz. the problem of explaining why our beliefs and desires have the content that they do. And while one could insist that this is not a problem that needs to be solved, it is only reasonable that the expressivist be allowed to assume otherwise. Recall the status of the dialectic thus far: the expressivist offers a way of assigning disjoint contents to the attitudes expressed by (1) and (2). The descriptivist complains that such an assignment of content is merely a stipulation: she requests an explanation of why those attitudes should be assigned the contents the expressivist assigns to them. It is only reasonable that the descriptivist herself has an explanation of why the state I am in when accepting (1) gets assigned the content that it does.

---

that, given our best shot at understanding what representation *is*, it follows that moral judgments are not representational.) But this is not relevant for our purposes.

26 Cf. Stalnaker 1984. Incidentally, some expressivists—e.g. Gibbard 1990, p. 109ff—seem to explicitly endorse a picture along these lines.

27 I suspect that much of what I have to say can be said even if one has alternative conceptions of what beliefs and desires are. But I want to work with a specific proposal, at least for the sake of concreteness.

28 Cf. the Postscript to Field 1978, in Field 2001. For critical discussion of the use of this analogy in this context, see Crane 1990.

Now, the story I've thus far described is overly simplistic and highly schematic. In particular, this approach allows for far too much indeterminacy. Since my beliefs and desires get their content *in tandem*, as it were, by changing the contents of my beliefs and desires appropriately, one can end up with distinct pairs of propositions that can characterize the same set of dispositions to act.[29] Thus, more often than not, versions of this story add an extra layer of complexity— say, by requiring that the relevant objects can be used to identify the causal role of the relevant mental state when taking into account an agent's environment.[30] At this level of detail, however, we can already draw some conclusions. And these conclusions are enough to raise doubts about PRIORITY OF CONTENT.

Consider the following analogy. Suppose you want an explanation for why it is colder in Boston than it is in Canberra.[31] I point out it is 20 degrees in Boston, that it is 85 degrees in Canberra, and that 20 is less than 85. I take it this would not be a satisfying explanation. The reason is that part of what justifies our assignment of numbers to magnitudes has to do with structural similarities between, on the one hand, the relations between numbers and, on the other, the relations between the magnitudes themselves. In other words, we use 20 to index the temperature in Boston and 85 to measure the temperature in Canberra among other things because 20 is smaller than 85 and it is colder in Boston than it is in Canberra.

Similarly, we assign some set-theoretic objects as the contents of some mental states because we think that the relations between those objects are structurally similar to those that obtain between the mental states themselves. If we want an explanation for why two mental states relate to each other in a particular way, pointing out that the corresponding set-theoretic objects relate to one another in the relevant way is not going to help. An explanation of the incompatibility of two mental states in terms of their truth-conditions is no more satisfying than the explanation of why Boston is colder than Canberra in terms of relations between numbers.

I should note that this is all compatible with the claim that facts about what content mental states have can do significant explanatory work.[32] The point here

---

29 We can think of this picture as one where representational contents are *models* of the relations among mental states, as does Egan 1992, p. 452f. The indeterminacy arises for the same sort of reasons that theories have unintended models. For an example, see Stalnaker 1984, p. 17ff.

30 See the discussion of information-theoretic approaches to intentionality below for an example of ways of doing this.

31 Importantly, you are not interested in a reason *to think* that it is colder in Boston than it is in Canberra. Stipulate you know that it is colder in Boston than it is in Canberra—you just traveled from one to the other, say. What you want is to know *why* this is so.

32 Cf. Jackson & Pettit 1988, Dretske 1988. See also Egan 2010: "[C]ontent ascription plays a crucial *explanatory* role: it is necessary to explain how the operation of a mathematically characterized process constitutes the exercise of a cognitive capacity in the environment in which the process is normally deployed. The device would compute the same mathematical function in any environment, but only in some environments would its doing so enable the organism to see." (p. 256)

is that facts about what content states have cannot be used to explain why states bear the sorts of relations to one another that are in turn used to explain why those states have the content that they have.

I find the above argument persuasive, but some might reasonably object. They will point out that I have ignored a possible way things could turn out, once the explanatory project comes to an end. Perhaps I am right that some facts about how states relate to one another are prior to facts about what content states have. But this does not mean that all such facts are. In particular, this does not mean that *inconsistency* facts are prior to what contents states have. Take again the temperatures of Boston and Canberra. There are facts about how Boston and Canberra relate to one another (at a particular time) that explain why we use 20 to measure the temperature of Boston and 85 to measure the temperature of Canberra. And surely we cannot explain any of *those* facts in terms of the relations between 20 and 85. But there may be *other* facts about how Boston and Canberra relate to one another that we can explain in terms of the relations between 20 and 85. In other words, it is compatible with all I have said that inconsistency facts are not among those that explain what contents mental states have.

3.3 Against Priority of Content, II

Fair enough. There is a gap in the argument that needs to be closed. Things *could* turn out so that inconsistency is explained by facts about contents, which in turn are explained in terms of facts about relations among mental states, even though the latter cannot by themselves explain inconsistency (although it is not obvious how that story would go). But I think we can close that gap if we look more closely at what an explanation of content is going to look like.

I noted above that if contents were assigned simply to index relations that obtain among mental states, content facts would be massively underdetermined. We can use different scales to index the temperatures of Boston and Canberra— relations among Boston and Canberra and their meteorological properties don't pick out one unique scale over another. Why isn't it the same with beliefs? Why not index the relations that obtain among my belief states with propositions other than those we would take to be the *representational* contents of those beliefs?[33]

The answer is that what contents beliefs have depends on more than just how beliefs relate to one another. As proponents of *information-theoretic* accounts of intentionality have long insisted, what representational content a mental state has essentially depends on what *information* is carried by that mental state.[34] Here, 'information' is a technical term, usually cashed out in terms of *indication*,

---

33 This is essentially the worry raised in Putnam 1980. Cf. also Crane 1990.
34 E.g. Dretske 1981, Fodor 1975, Millikan 1984. As Fodor put it, "as far as I can see, of the various proposals around for a naturalistic account of content, only the informational one appears to have a prayer of working." Fodor 1994, p. 6

tracking, or *causal co-variation*. On this picture, a belief has representational content in virtue of indicating how things stand with the world. For example, my belief that Tom is a cannibal—understood as a particular physical state I am in—has the content that it does partly because it is a state that is typically caused by the fact that Tom is a cannibal. In 'normal conditions', one can infer that Tom is a cannibal from the fact that I believe that Tom is a cannibal.

The hard work for these views begins when trying to give an account of what 'normal conditions' are so that if (and only if) we take an agent's state s to have content p, then in normal conditions, the agent is in s if and only if p is true. For if such an account is to help provide an explanation of content, it must be given in non-intentional terms—in particular, we cannot presuppose anything about the representational properties of s. And while it is plausible that we can give such an account so that it works for simple, perceptual beliefs, it is not obvious that we can give one that works in full generality.

Let me illustrate this point with an example. Consider a belief of an agent, Lars, which we take to represent that there is an elephant in front of him. Call that belief state e. Sometimes, Lars is in e when in fact there is an elephant in front of him. Sometimes, however, he is in e when there is no elephant in the vicinity. Lars is somewhat fond of recreational drugs, and whenever he is under the effect of a hallucinogenic, seeing a picture of an elephant causes him to be in state e. We can conclude that Lars is in state e if and only if either there is an elephant in front of him or there is a picture of an elephant in front of him. If the representational content of e were just the state of the world that co-varies with Lars being in state e, then the content of e would not be that there is an elephant in front of him. Rather, it would be that there is an elephant or a picture of an elephant in front of him.

This is where talk of normal conditions comes in handy. Plausibly, a situation in which Lars is under the influence of hallucinogenic drugs is not normal: there is something about Lars' cognitive or perceptual system that is off, and this because of the effect of the drugs. And plausibly, we can say what it is about the system in virtue of which it is malfunctioning in purely non-intentional terms. So while Lars being in e reliably co-varies with there being an elephant or a picture of an elephant in front of him, *in normal conditions* his being in state e does not. Instead, in normal conditions, his being in state e co-varies precisely with there being an elephant in front of him. To put it in slightly different terms: we can rule out some cause c of e from being content-determining by saying that c only causes e when normal conditions do not obtain.

But now suppose Lars has a tendency for drawing odd inferences.[35] Whenever he believes that there is a dog in the house, he comes to believe that there is a dog and a cat in the house; whenever he believes that there is a cat in the house, he comes to believe that there is a dog and a cat in the house. Consider

35 Cf. Field 1990, p. 108f.

now his belief state that there is a dog and a cat in the house, which we can label D+C. Given the description of the setup, D+C does not reliably co-vary with there being dogs and cats in the house. Rather, it reliably co-varies with there being dogs *or* cats in the house. So why isn't the content of D+C that there are dogs or cats in the house?[36]

One would hope that, here again, appeal to normal conditions would do the trick. Yet it is hard to see how that would work—at least if we are supposed to cash out normality in non-intentional terms. It seems plausible that Lars' perceptual machinery is working properly—he is not having hallucinations of any kind, nor is he under the effect of some drug. Intuitively, what seems to be going on here is that Lars is not reasoning according to the rules of logic. It is his *logical capacities* that are malfunctioning in this context. And if this is to help us explain why D+C has the content we take it to have, there must be some way of characterizing what it is for Lars' logical capacities to function properly in non-intentional terms.[37]

We can put the point in slightly different terms. Suppose we could impose a constraint on the assignment of content to Lars' mental states of the following sort:

> CONSTRAINT: Whenever state *s* is inconsistent with state *s′*, the content assigned to *s* must be incompatible with the content assigned to *s′*.

Then we could rule out the assignment of content that assigns a disjunction to D+C by saying that it fails to satisfy CONSTRAINT. After all, the state Lars is in when he believes that there are cats and dogs in the house is inconsistent with the state he is in when he believes that there are no cats in the house. And the disjunctive content *that there are cats or dogs in the house* is not inconsistent with the content *that there are no cats in the house*.

Of course, if we are to appeal to CONSTRAINT as a way of solving this problem, we need to reject PRIORITY OF CONTENT. So unless the descriptivist has an alternative solution to offer, she will have to reject PRIORITY OF CONTENT.

Now, you may not find this example especially compelling. For, you might reasonably think, Lars' doxastic dispositions make him borderline unintelligi-

---

36 As the example suggests, this problem is related to the so-called disjunction problem—cf. Fodor 1987, 1990. The two problems are distinct, however, in that a solution to Lars's problem would not amount to a solution to the disjunction problem.

37 Cf. Fodor 1994, p. 76f: "Semantic atomism is the idea that the meaning of your words—mutatis mutandis, the contents of your thoughts—is metaphysically independent of the inferences you are preferred to draw. […] I think that the puzzles about inscrutability show that semantic atomism is probably false in this strong form. The ontology of a language supervenes not on mind/world connections alone, but on mind/world connections *plus logical syntax*. [W]hat [this] tells us about naturalism in semantics is something we already knew; viz. that the program fails unless there is a naturalistic account of inferring." I should add here, however, that Fodor may have well changed his mind about some of this—see Rey 2004

ble.[38] Perhaps it is just not possible to have the dispositions I stipulate in this case. If so, I haven't given any reason to think that CONSTRAINT should play a role in explaining why some thoughts have the content that they do. But a slight variant of this case might do the trick instead.[39]

Suppose Lars believes, falsely, that cats and dogs always follow each other around. Wherever you see a dog, there is a cat somewhere in the vicinity, and vice versa. So, whenever Lars comes to believe there's a dog in the house, he forms the belief that there is a dog in the house, he comes to be in D+C—similarly, whenever he comes to believe there's a cat in the house, he comes to be in D+C. Now, surely, D+C reliably co-varies with there being a dog *or* a cat in the house. And it does not reliably co-vary with there being a dog and a cat in the house. So we need some story as to why D+C gets to have the content that it does, viz. that there is a dog and a cat in the house.

Again, we might try saying that in this case, what matters is whether D+C co-varies *in normal conditions* with there being a cat and a house. And presumably what we'd want to say is that normal conditions obtain if and when Lars does not have the false belief that cats and dogs always follow each other around. But this alone would not do. For we need to specify what normal conditions are in non-intentional terms.

Here too appealing to CONSTRAINT could help solve our little problem. But in appealing to CONSTRAINT, we would have to give up on PRIORITY OF CONTENT.

What would an alternative solution look like? One might insist that there are logical relations among mental states *other than* inconsistency that play a role in determining what contents those states have. For instance, one might replace CONSTRAINT with:

> CONSTRAINT*: The content assigned to *s* must entail the one assigned to *s'* only when being in state *s* entitles an agent to being in state *s'*.

If our assignment of content is to satisfy CONSTRAINT*, we could not assign a disjunction to D+C. For Lars's belief that there are cats in the room does not entitle him to being in D+C. Indeed, his logical failure lies precisely in the fact that he moves from a state he is entitled to be to a state he is not entitled to be. Assigning to D+C the disjunctive content—which is entailed by the content of his belief that there are cats—would run afoul of CONSTRAINT*.

Now, strictly speaking, CONSTRAINT* is not incompatible with PRIORITY OF CONTENT. It is only incompatible with a related principle, viz. that entitlement relations among beliefs obtain in virtue of the contents that those states have. And while one could hold on to this related principle while denying PRIORITY OF CONTENT, it is hard to see what the motivation for doing so would

---

38 Thanks to an anonymous reviewer for pushing me on this point.
39 This version of the case is much closer to the one given in Field 1990.

be.[40] After all, the likely explanation for why basic entitlement relations are not explained in terms of the truth-conditional contents of the relevant states is that broadly logical relations among mental states are prior to whatever contents they have.

Admittedly, I have not argued that there is no alternative solution to this problem. For all I have said, there could be a way of explaining content that avoids having to postulate logical relations among mental states that are prior to facts about the contents of mental states.[41] But at the very least, the burden should now be on the descriptivist to give a defense of PRIORITY OF CONTENT in light of these objections.

4 Conclusions

The negation problem is, at bottom, an explanatory challenge: that of explaining NEGATION. The presumption is that descriptivists, but not expressivists, have a good explanation of NEGATION. I have argued that this presumption is mistaken. To the extent that expressivists are unable to explain NEGATION, so are descriptivists. So if NEGATION is something that needs to be explained, then this is a problem we all have.

My argument was based on two main claims. First, that for DESCRIPTIVIST EXPLANATION to be a better explanation than EXPRESSIVIST EXPLANA-TION, it must be that PRIORITY OF CONTENT is true. Second, that if some of the more promising attempts at explaining of content—at solving the problem of intentionality—are on the right track, PRIORITY OF CONTENT must be false. Thus, unless we give up on the project of explaining content, we have good reason to reject PRIORITY OF CONTENT, and thus that DESCRIPTIVIST EXPLANATION is any better than EXPRESSIVIST EXPLANATION. This is not to say that either of those explanations is any good—I want to remain neutral on that. My point has been to argue that if expressivists have a problem with negation—if they have no adequate explanation of NEGATION—then it is a problem we all have.

---

40 Interestingly, there are good reasons for thinking that, in order to explain why negation means what it does, *while insisting that negation has its 'classical' meaning*, we need to rely on some prior attitude of rejection or, perhaps, disagreement. See e.g. Price 1990, Rumfitt 2000, Price n.d.

41 It is worth noting, however, that the view that the meaning of logical vocabulary should not be explained in terms of representational notions is not especially controversial. Cf. e.g. Boghossian 2011, p. 493: "it's hard to see what else could constitute meaning *conjunction* by 'and' except being prepared to use it according to some rules and not others (most plausibly, the standard introduction and elimination rules for 'and'). Accounts that might be thought to have a chance of success with other words—information-theoretic accounts, for example, or explicit definitions, or teleological accounts—don't seem to have any purchase in the case of the logical constants." Even Jerry Fodor, the arch-enemy of any inferentialist theory of meaning, ended up granting that the meaning of logical constants cannot explained in representational terms, but rather in purely inferential terms. See Fodor 1990, p. 110-111 and Fodor 1994, p. 74-77.

In this respect, my solution to the negation problem is more ambitious than a recent one offered by Allan Gibbard.[42] Gibbard argues that if descriptivists are allowed to rely on something like PRIORITY OF CONTENT, the expressivist is allowed to rely on a different claim of explanatory priority: they are allowed to rely on the claim that facts about inconsistency are prior to facts about what contents states have. As I understand it, his claim is that both claims of explanatory priority are on a par. If I am right, however, this is not so: if we are to arrive at an adequate explanation of content, I have urged, we must reject PRIORITY OF CONTENT. Facts about how our mental states logically relate to one another are prior to facts about what representational content, if any, those states happen to have.[43]

References

Baker, Derek & Jack Woods. 2015. How Expressivists Can and Should Explain Inconsistency. *Ethics* 125(2). 391–424.

Blackburn, Simon. 1990. Hume and Thick Connexions. *Philosophy and Phenomenological Research* 50. 237–250.

Blackburn, Simon. 1998. *Ruling Passions*. Oxford University Press.

Boghossian, Paul A. 2011. Williamson on the a Priori and the Analytic. *Philosophy and Phenomenological Research* 82(2). 488–497.

Crane, Tim. 1990. An Alleged Analogy Between Numbers and Propositions. *Analysis* 50(4). 224–230.

Davidson, Donald. 1974. Belief and the Basis of Meaning. English. *Synthese* 27(3-4). 309–323.

Davis, Wayne A. 2002. *Meaning, Expression, and Thought*. Cambridge: Cambridge University Press.

Dreier, James. 1999. Transforming Expressivism. *Noûs* 33(4). 558–572.

Dreier, James. 2004. Meta-Ethics and the Problem of Creeping Minimalism. *Philosophical Perspectives* 18(1). 23–44.

Dreier, James. 2006a. Disagreeing (About) What to Do: Negation and Completeness in Gibbard's Norm-Expressivism. *Philosophy and Phenomenological Research* 72(3). 714–721.

---

42 Gibbard 2012, Appendix 2.

43 An interesting upshot of this conclusion is that that some well-known arguments for the view that intentions involve belief—what Michael Bratman calls *cognitivism about practical reason*—are undermined, for they presuppose that norms of epistemic rationality can be explained in terms of their truth-conditions (Cf. Wallace 2001). Very roughly, the idea is that in order to explain why there are consistency requirements on *intention*, we must assume that intentions involve beliefs. But the reason this is supposed to help explain why it is irrational to intend to ϕ while at the same intending to not-ϕ is that the beliefs that are thereby implicated have incompatible truth-conditions, and that it is irrational to have beliefs with incompatible truth-conditions. If I am right, however, the putative explanandum is no more in need of an explanation than the explanans.

Dreier, James. 2006b. Negation for Expressivists: A Collection of Problems With a Suggestion for Their Solution. In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 1, 217–233. Oxford University Press.

Dreier, James. 2009. Relativism (And Expressivism) and the Problem of Disagreement. *Philosophical Perspectives* 23(1). 79–110.

Dretske, Fred. 1981. *Knowledge and the Flow of Information*. Cambridge, Mass.: MIT Press.

Dretske, Fred. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, Mass.: MIT Press.

Dummett, Michael. 1991. The Relative Priority of Thought and Language. In. *Frege and Other Philosophers*. Oxford University Press. 315–324.

Egan, Frances. 1992. Individualism, Computation, and Perceptual Content. *Mind* 101(403). 443–459.

Egan, Frances. 2010. Computational Models: A Modest Role for Content. *Studies in History and Philosophy of Science* 41(3). 253–259.

Field, Hartry. 1978. Mental Representation. *Erkenntnis* 13(1). 9–61.

Field, Hartry. 1990. 'Narrow' Aspects of Intentionality and the Information-Theoretic Approach to Content. In Enrique Villanueva (ed.), *Information, Semantics and Epistemology*, 102–116. Oxford: Basil Blackwell.

Field, Hartry. 2001. *Truth and the Absence of Fact*. New York: Oxford University Press.

Field, Hartry. 2009. Epistemology Without Metaphysics. *Philosophical Studies* 143(2). 249–290.

Fodor, Jerry A. 1975. *The Language of Thought*. Cambridge, Mass.: Harvard Univ Press.

Fodor, Jerry A. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press.

Fodor, Jerry A. 1990. *A Theory of Content and Other Essays*. Cambridge, Mass.: MIT Press.

Fodor, Jerry A. 1994. *The Elm and the Expert*. Cambridge, Mass.: MIT Press.

Frege, Gottlob. 1893. *Grundgesetze Der Aritmetik*. Vol. 1. Jena. Translated as Gottlob Frege. 1967. *Basic Laws of Arithmetic*. Trans. by Montgomery Fush. Berkeley, CA: University of California Press. Page numbers refer to this translation.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, Mass.: Harvard University Press.

Gibbard, Allan. 1994. Meaning and Normativity. *Philosophical Issues* 5. 95–115.

Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge, Mass.: Harvard University Press.

Gibbard, Allan. 2012. *Meaning and Normativity*. Oxford: Oxford University Press.

Grice, H. P. 1957. Meaning. *Philosophical Review* 66(3). 377–388.

Grice, H. P. 1969. Utterer's Meaning and Intention. *Philosophical Review* 78(2). 147–177.

Horwich, Paul. 1994. The Essence of Expressivism. *Analysis* 54(1). 19–20.

Incurvati, Luca & Julian J. Schlöder. 2019. Inferential Expressivism and the Negation Problem. *Oxford Studies in Metaethics*. Forthcoming.

Jackson, Frank & Philip Pettit. 1988. Functionalism and Broad Content. *Mind* 96(387). 381–400.

Lewis, David. 1975. Languages and Language. *Minnesota Studies in the Philosophy of Science* 7. 3–35. Reprinted in Lewis 1983, pp. 163–189.

Lewis, David. 1983. *Philosophical Papers*. Vol. 1. New York: Oxford University Press.

McHugh, Conor & Daniel Whiting. 2014. The Normativity of Belief. *Analysis* 74(4). 698–713.

Millikan, Ruth G. 1984. *Language, Thought and Other Biological Objects*. Cambridge, Mass.: MIT Press.

Ninan, Dilip. 2010. De Se Attitudes: Ascription and Communication. *Philosophy Compass* 5(7). 551–567.

Pérez Carballo, Alejandro. 2014. Structuring Logical Space. *Philosophy and Phenomenological Research*. Early View.

Price, Huw. n.d. 'Not' Again. Unpublished manuscript, Cambridge University.

Price, Huw. 1983. Does 'Probably' Modify Sense? *Australasian Journal of Philosophy* 61(4). 396–408.

Price, Huw. 1990. Why 'Not'? *Mind* 99(394). 221–238.

Putnam, Hilary. 1980. Models and Reality. *The Journal of Symbolic Logic* 45(3). 464–482.

Restall, Greg. 2005. Multiple Conclusions. In Luis Valdes-Villanueva Petr Hajek & Dag Westerstahl (eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the Twelfth International Congress*, 189–205. London: Kings' College Publications.

Rey, Georges. 2004. Fodor's Ingratitude and Change of Heart? *Mind & Language* 19(1). 70–84.

Rosen, Gideon. 1997. Who Makes the Rules Around Here? *Philosophy and Phenomenological Research* 57(1). 163–171.

Rosen, Gideon. 2001. Brandom on Modality, Normativity and Intentionality. *Philosophy and Phenomenological Research* 63(3). 611–623.

Rosen, Gideon. 2010. Metaphysical Dependence: Grounding and Reduction. In Bob Hale & Aviv Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press.

Rumfitt, Ian. 2000. "Yes" and "No". *Mind* 109(436). 781–823.

Schiffer, S. 1982. Intention-Based Semantics. *Notre Dame Journal of Formal Logic* 23(2). 119–156.

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

Schroeder, Mark. 2008b. Expression for Expressivists. *Philosophy and Phenomenological Research* 76(1). 86–116.

Schroeder, Mark. 2008c. How Expressivists Can and Should Solve Their Problem With Negation. *Noûs* 42(4). 573–599.

Smiley, Timothy. 1996. Rejection. *Analysis* 56(1). 1–9.

Smith, Michael. 1994. Why Expressivists About Value Should Love Minimalism About Truth. *Analysis* 54(1). 1–11.

Stalnaker, Robert C. 1984. *Inquiry*. Cambridge, Mass.: Bradford books, MIT Press.

Stevenson, Charles Leslie. 1937. The Emotive Meaning of Ethical Terms. *Mind* 46(181). 14–31.

Unwin, Nicholas. 1999. Quasi-Realism, Negation and the Frege-Geach Problem. *The Philosophical Quarterly* 49(196). 337–352.

Unwin, Nicholas. 2001. Norms and Negation: A Problem for Gibbard's Logic. *The Philosophical Quarterly* 51(202). 60–75.

van Roojen, Mark. 1996. Expressivism and Irrationality. *Philosophical Review* 105(3). 311–335.

Wallace, R. Jay. 2001. Normativity, Commitment, and Instrumental Reason. *Philosophers' Imprint* 1(4). 1–26.

Yalcin, Seth. 2011. Nonfactualism About Epistemic Modality. In Andy Egan & Brian Weatherson (eds.), *Epistemic Modality*, 295–334. Oxford: Oxford University Press.