

# Expression for Expressivists

MARK SCHROEDER

*University of Southern California*

Expressivism's central idea is that normative sentences bear the *same* relation to non-cognitive attitudes that ordinary descriptive sentences bear to beliefs: the *expression* relation. Allan Gibbard tells us that "that words express judgments will be accepted by almost everyone" – the distinctive contribution of expressivism, his claim goes, is only a view about what *kind* of judgments words express. But not every account of the *expression* relation is equally suitable for the expressivist's purposes. In fact, what I argue in this paper, considering four possible accounts of *expression*, is that how suitable each account is for the expressivist's purpose varies in proportion to how *controversial* it is. So Gibbard is wrong – if *expression* is to get expressivism off the ground, then it will be enormously controversial whether words *do* express judgments. And thus expressivism is committed to strong claims about the semantics of *non*-normative language.

That words express judgments will, of course, be accepted by almost everyone.  
—Allan Gibbard<sup>1</sup>

If moral judgements cannot express beliefs, what do they express?  
—Alexander Miller<sup>2</sup>

## 1

### 1.1 Introduction

According to expressivism, "to make a normative judgment is to express a non-cognitive attitude." A great deal of ink has been spilled over what *kind* of non-cognitive attitude various normative judgments express, and over various problems raised for this kind of view. But very little has been said about what *expression* is, in the first place. If Allan Gibbard is right, then this should not be very surprising, for "[t]hat words express judgments will, of course, be accepted by almost everyone".<sup>3</sup> This suggests both that we all understand what is meant by

---

<sup>1</sup> Gibbard (1990, 84).

<sup>2</sup> Miller (2003, 6).

<sup>3</sup> Gibbard (1990, 84).

“express” and that the uncontroversial feature of expressivism is not that it holds sentences to express mental states, but what *kind* of mental states it holds them to express. And this view is presupposed by typical characterizations of expressivism in the literature.<sup>4</sup> But in this paper I will be arguing otherwise. Once we start to ask what this mysterious *expression* relation is, I will be arguing, the uncontroversial senses in which words express judgments will not do at all for the expressivists’ purposes. Even the moderately controversial account adopted by Allan Gibbard, I’ll suggest, is uncomfortably unstable. Expressivists, I’ll be arguing, need an account of *expression* that commits them to a quite radical view about the foundations of the semantic content of ordinary, descriptive sentences.

Distinguish two very different kinds of view which have historically come under the heading of expressivism: *pure* and *adulterated*. Both kinds of expressivists claim that it is part of the meaning of normative sentences to be used to express non-cognitive attitudes. But while *pure* expressivists hold that this exhausts their meaning, *adulterated* expressivists hold that normative sentences also have ordinary propositional contents.<sup>5</sup> I will *not* be concerned with adulterated expressivism in this paper. Adulterated expressivists *have* put some significant work into telling us what they mean by saying that normative utterances *express* non-cognitive attitudes.<sup>6</sup> But adulterated expressivism is less common, and with the exception of Allan Gibbard, *pure* expressivists have said very little about this, taking it for granted that we all know what they mean by “express.” And even Gibbard, when he outlines his view about what *expression* is, emphasizes that it is supposed to be a view

---

<sup>4</sup> For example, Richard Joyce writes that “[n]on-cognitivism is the metaethical view according to which public moral judgements do not express beliefs, in spite of the fact that they are formed in the indicative mood. [...] [E]xpressivism—holds that moral judgements function to express desires, emotions, or pro/con attitudes” (2002, 336-337). As Gibbard explains in *Thinking How to Live*, to expressivist project is to “[a]sk what *states of mind* ethical statements express... The term ‘expressivism’ I mean to cover any account of meanings that follows this indirect path: to explain the meaning of a term, explain what states of mind the term is used to express” (2003, 6-7).

<sup>5</sup> Some adulterated expressivists say this in order to capture a sense in which normative sentences can be true or false, some in order to solve the Frege-Geach problem (Alm (2000), Ridge (2006)), some in order to allow for normative realism (Copp (2001)), and some simply in order to answer our question—to say what *expression* is, which distinguishes it from *reporting* (Barker, (2000)). Stevenson seems to have been some kind of adulterated expressivist (Stevenson (1937), (1944)). Obviously, some of the standard arguments for expressivism, for example the argument from motivational internalism, would seem at best to support adulterated expressivism, unless further supplemented. On whether adulterated expressivism has hopes of solving the Frege-Geach problem, see van Roojen (2005), Schroeder (ms).

<sup>6</sup> See Stevenson (1937 & 1944), Copp (2001), and Barker (2002).

about something that we can all accept: that “words express judgments”. So from here forward, by “expressivism” I will mean *pure* expressivism. The theories of pure expressivists place the most weight on the notion of *expression*, but they have told us the least about it. It is my task in this paper to investigate what they *can* and *should* say, and *why*.

The structure of the paper will be straightforward. In parts 1 and 2 I will argue that we should understand pure expressivism as committed to two further theses. First, in parts 1 and 2, that normative sentences bear the *same* “expression” relation to non-cognitive attitudes that ordinary descriptive sentences bear to beliefs. And second, in the remainder of part 2, that ordinary descriptive sentences, with ordinary propositional contents, *inherit* those contents from the beliefs that they express. I’ll use these two commitments to evaluate various candidates for the *expression* relation in the remainder of the paper. In part 3 I will survey the three most obvious candidates for the *expression* relation: the *shared-content* account, the *causal* account, and the *implicature* account. All three are fairly obviously unsuitable for an expressivism with the commitments argued for in parts 1 and 2. In part 4 I consider a fourth account of the *expression* relation due to Allan Gibbard, *indicatory* expressivism, which is well-designed to solve the problems for the other accounts. But I argue that the *way* that it solves these problems commits it to more than is necessary in order to solve them, and as a result makes indicatory expressivism uncomfortably unstable.

Finally, in part 5 I introduce the star of our story, *assertability* expressivism. Assertability expressivism is designed to get around all of the problems raised for other accounts of the *expression* relation in a way that is minimally committal, so it is the version expressivists should adopt. I’ll be arguing that if this is what expressivists do and should mean by “express,” then expressivism is *not* just a distinctive view about normative thought, as Gibbard and Horgan and Timmons would sometimes have us to believe.<sup>7</sup> It is a radical view about the foundations of semantic content—including for ordinary descriptive sentences. I don’t claim that this is a *problem* for assertability expressivism. I just claim that it is something that we need to understand, if we want to properly *assess* pure expressivism. Pure expressivism can only be as palatable and as well-motivated as this radical view of the foundations of ordinary descriptive content.

---

<sup>7</sup> “Expressivism in ethics is a metaethical view according to which (roughly) a typical moral judgment functions to express some psychological state other than a descriptive belief, such as some desire, intention, or other motivating state” (2006, 73).

## 1.2 The Modal and Disagreement Problems

In the remainder of part I I will be arguing that pure expressivism needs to be understood as committed to holding that normative sentences bear the *same* relation to non-cognitive attitudes as ordinary descriptive sentences bear to ordinary propositional beliefs. I will call this the *parity* thesis. There are three sources of pressure in this direction—from weakest to strongest, these derive from the attractions of pure expressivism vis-à-vis other non-cognitivist views, from the source of the advantages of pure expressivism over Cognitivist Speaker Subjectivism—the view that normative sentences *report* non-cognitive mental states of the speaker—and from the form of the solution to the Frege-Geach problem that is both relatively standard and apparently the only one available to pure expressivism.

The first of these sources of pressure is easy to understand. If Gibbard is right and “[t]hat words express judgments will, of course, be accepted by almost everyone,”<sup>8</sup> then *all it takes* to provide a non-cognitivist account of normative discourse is to take a non-cognitivist view about *what kind* of judgment is expressed by normative words. There will be no complicated rigamarole giving separate accounts of normative speech acts and of normative thought—*thinking* that murder is wrong is clearly not issuing a command or exhortation, so what *is* it, on accounts that tell us this is the role of normative “assertions”? If the parity thesis is right, then a non-cognitivist account of normative *language* will *fall out* of a non-cognitivist account of normative thought. We will get our accounts of normative language and normative thought in one swoop. And this has been one of the chief perceived attractions of pure expressivism ever since Gibbard published *Wise Choices, Apt Feelings*.<sup>9</sup>

This thought is shared, I think, by anyone who writes about expressivism by saying that according to expressivists, “moral judgments function to express some psychological state other than a descriptive belief.”<sup>10</sup> The use of such a sentence implies that there is a single *expression* relation at stake, and buys into the Gibbardian idea that we should understand expressivism as holding the *expression* relation fixed, but substituting a different account of normative thought. It is also the thought that will turn out to be wrong, if even a significant part of what I argue in this paper is correct. But this source of pressure to allow that there is only one *expression* relation at stake in both

---

<sup>8</sup> Gibbard (1990, 84).

<sup>9</sup> Gibbard, (1990). See also Blackburn (1998).

<sup>10</sup> Horgan and Timmons (2006, 73).

normative and ordinary descriptive sentences is relatively mild. The next two sources of pressure, however, are I think quite significant.

To understand the second source of pressure to accept the parity thesis, we have to appreciate why expressivism is supposed to be an improvement on Cognitivist Speaker Subjectivism, in the first place. Cognitivist Speaker Subjectivism (CSS, from here forward) is exemplified by the view that “murder is wrong” is true relative to a context of utterance *C* just in case “I disapprove of murder” is true relative to *C*.<sup>11</sup> So it is committed to holding that “wrong,” and other normative terms, are indexicals. But then it follows from some straightforward properties of the way that indexicals function in embedded contexts that CSS is up against some important difficulties. Both have a lot to do with the accusation that CSS amounts to an objectionable kind of *relativism*. It is the elegant way in which pure expressivism can avoid these difficulties that constitutes the second motivation for the parity thesis.

First is the *modal* problem. If “murder is wrong” is true relative to *C* just in case “I disapprove of murder” is, then the same goes for “if I didn’t disapprove of murder, then it wouldn’t be wrong” and “if I didn’t disapprove of murder, then I wouldn’t disapprove of murder.” Since the latter is indexically valid, in Kaplan’s sense,<sup>12</sup> so is the former. And so it is true in the mouth of any speaker. But that’s wrong—no one has that kind of control over what is right and wrong.

Second is the *hard* version of the *disagreement* problem. I call it that because it arises when two people are in disagreement—for example, when Sally sincerely asserts, “murder is wrong” and Jim sincerely asserts, “murder is not wrong.” But it is important to distinguish it from *shallow* conceptions of the disagreement problem. On a *shallow* conception of the disagreement problem, the problem is just that since “wrong” is an indexical, Sally and Jim can’t really be *disagreeing* when they say these things. There is a standard expressivist response to this objection—that approving of murder and tolerating it *are* in a kind of disagreement—in *attitude*.<sup>13</sup> And this response can be adopted by CSS. But passing over whether this is an adequate response, the *hard* problem of disagreement is that even if Sally *can* count as disagreeing with Jim by saying, “murder is wrong,” she *can’t* disagree with him by

---

<sup>11</sup> Nothing says that either expressivists or Cognitive Speaker Subjectivists have to think that there is any English word for the non-cognitive attitude that is expressed or reported by “murder is wrong.” But I’ll use “disapprove” as a stipulative term for whatever attitude it is that such theorists have in mind. Nothing will turn on this for my purposes in this paper.

<sup>12</sup> Kaplan (1977).

<sup>13</sup> The response dates to Stevenson (1937).

saying that what he says is *false*. For on the contrary, *even if* Sally and Jim count as disagreeing, each must still allow that what the other says is *true*.<sup>14</sup> But that is wrong. Sally should be able to disagree with Jim just as well by saying that what he said is false as by negating it.

The important thing to observe about the modal and disagreement problems is that there is nothing significant about CSS that gives rise to these problems. Importantly, we would get *exactly* the same problems if we conflated “grass is green” with “I believe that grass is green.” If “grass is green” were true relative to *C* just in case “I believe that grass is green” was, then the same would follow for “if I didn’t believe that grass is green, then grass wouldn’t be green” and “if I didn’t believe that grass is green, then I wouldn’t believe that grass is green.” Since the latter is indexically valid, so is the former. But no one has that kind of control over the color of grass. Similarly, if Rob and Carla disagree about the color of grass, and Rob sincerely asserts, “grass is not green,” it follows from this conflation that even Carla must acknowledge that what Rob said is true—so long as Rob does *believe* that grass is not green—even if she can go on to insist, “grass is *so* green.”

### 1.3 The Expression Relation

The key idea of pure expressivism—indeed, its fundamental insight, if it really is an insight—is that these are the *same problems* as the problems for CSS. And so the elegant solution of pure expressivism is to propose that they issue from the *same mistake*. CSS, the claim goes, is making the same mistake as we would make if we confused “grass is green” with “I believe that grass is green.” Though “murder is wrong” is intimately *connected* with “I disapprove of murder,” it is not by sharing truth-conditions. It is by bearing the *same* relation to it as “grass is green” bears to “I believe that grass is green”—whatever that is. It is by *expressing* the mental state that the other sentence *reports*.

The central idea, here, is not simply a diagnosis of what went wrong with CSS—although I think that is the way it is usually understood. If all that we were doing was diagnosing what went wrong with CSS, that wouldn’t commit us to any view about the relation between “murder is wrong” and disapproval of murder—we could just insist that it is *not* what CSS says it is. But that would only solve one half of the problem. It would establish that we were not making the *same* mistake as CSS, but it wouldn’t constitute any kind of evidence that we were not going to be subject to some extended version of the problem. And

---

<sup>14</sup> Compare Cappelen and Lepore (2005).

importantly, it is a persistent objection to even expressivism that it seems to amount to a kind of “relativism”—making the wrongness of murder somehow *depend* on attitudes in something like the way that the modal and disagreement problems show that CSS makes it.<sup>15</sup> But the *pure* expressivist can do this solution one better. According to the key idea of pure expressivism, these problems *couldn't* arise again for expressivism. Since “murder is wrong” is to “I disapprove of murder” in exactly the same way as “grass is green” is to “I believe that grass is green,” it follows that *no matter what* explains why there are no modal or disagreement problems for “grass is green” will explain why there are none for “murder is wrong.”

The brilliance of this idea lies in the fact that this solution seems, at least at first glance, to follow *no matter what* the *expression* relation turns out to be. If the parity thesis is true, then it seems that we don't even have to know *what* the *expression* relation is. It follows that there are no new problems raised by normative language, over and above the problems raised by normative thought. This, I think, is the Big Idea behind the entire research program of expressivism. It is the whole source of the idea that pure expressivism is more attractive than CSS. And it explains—importantly!—why pure expressivists have not, by and large, felt any burden to tell us what *expression* is.<sup>16</sup> Their central idea is that we already know.

## 2

### 2.1 Frege-Geach and Composition

This should be enough, I think, to persuade us that pure expressivists are committed to the parity thesis. It is at least enough to indicate that there is enough pressure for pure expressivists to accept the parity thesis that it would be worthwhile to investigate the consequences of this view. But there is one more source of pressure to accept the parity thesis which is, I think, the clincher. It is that the parity thesis is a consequence of the standard expressivist solution to the Frege-Geach problem, and this standard solution is the only one available to pure expressivists. The problem has to do with how normative terms function in embedded contexts, and is usually set up by considering an example like the following argument:

---

<sup>15</sup> See, for example, Zangwill (1994) and Shafer-Landau (2003, 30-33).

<sup>16</sup> Again, with the minor exceptions of a brief discussion in Gibbard (1990, 84-86) and again in Gibbard (2003, 75-79)—the latter of which is really mostly negative, not really adding anything to the positive account in Gibbard (1990). See section 4.1 for further discussion.

- 1 If being friendly is wrong, then being friendly to strangers is wrong.
- 2 Being friendly is wrong.
- 3 Being friendly to strangers is wrong.

This argument is valid. And so we *know* that “being friendly is wrong” must mean the same thing in the antecedent of **1** as in **2**. But it *seems* (and we will come back to re-consider this later, in section 3.2) that one can accept or assert **1** without expressing the same attitude that would be expressed by an assertion of **2**.<sup>17</sup>

So the problem has often been described as that of explaining why arguments like this are valid, if “being friendly is wrong” does not express the same attitude in **1** as in **2**. Adulterated expressivists sometimes seem to understand the problem in this way, for example. But the real problem is that of explaining why “being friendly is wrong” *means* the same thing in the antecedent of **1** as in **2**—the validity of the argument is primarily supposed to prove *that* it must. If it does not express the same attitude (it seems obvious that it does not, but I’ll question that later), and all there is to what it means is what attitude it expresses (as pure expressivism claims), then there is at least some puzzle about how it could mean the same thing in each context.

The key idea used to solve this problem by both Simon Blackburn and Allan Gibbard is to remind us of the Fregean idea that a word has a meaning only in the context of a proposition.<sup>18</sup> To see what “being friendly is wrong” means in the context of **1**, we have to see what contribution it makes to the meaning of the whole sentence. If it makes the *same contribution* to the meaning of the whole in **1** as it does in **2**, then it is being used in the same sense. This, after all, is exactly how truth-conditional semanticists see it. They don’t suppose that **1** can only be true if it is true that being friendly is wrong—they say that “being friendly is wrong” counts as meaning the same thing in **1** as in **2** because it makes the same contribution to the truth-conditions of the whole. The only problem is that ordinary truth-conditional semanticists explain why it makes the same contribution by giving the meaning of the conditional as a function from the propositions expressed by the antecedent and the consequent to the proposition expressed by the whole, and expressivists haven’t allowed that there are such things for “if...then” to be a function *of* in the case of normative sentences.

---

<sup>17</sup> Some version of this challenge was independently offered by Geach (1960 & 1965) and by Searle (1962 & 1969).

<sup>18</sup> Blackburn (1973, 1984, 1988 & 1998), Gibbard (1990 & 2003). “Since it is only in the context of a proposition that words have any meaning...” Frege (1950, 73).



So the Big Idea of pure expressivists about the Frege-Geach problem is to treat the logical connectives not as functions from propositions to propositions, but as functions from *mental states* to *mental states*. On this view, the way to solve the Frege-Geach problem is to explain the meaning of **1** by saying what mental state it expresses, as a function of the mental states expressed by the antecedent and by the consequent. Of course, this doesn't suffice at all to explain why the inference is valid—obviously even if we were doing ordinary truth-conditional semantics, not just *any* function filled in for “if...then” would make the inference valid. And so there expressivists have differed greatly from each other and from their past selves about just *how to characterize* the mental state expressed by **1**, and about why being in it and in the mental state expressed by **2** is “inconsistent” with not being in the mental state expressed by **3** in a sense that is sufficiently robust to make the inference count as “valid.” Indeed, there is a good question as to whether it is possible to do this.<sup>19</sup>

But this much suffices, nevertheless, to explain why “being friendly is wrong” means the same thing in the antecedent of **1** as self-standing, in **2**. And that is all that we need, for our purposes. We can see that this Big Idea is part and parcel of the Big Idea of pure expressivism full stop—that normative language should not raise any new problems over and above those raised by normative thought. Here, again, the Big Idea is to treat normative language in such a way that the problems about normative language—in this case, how this inference could be valid—reduce to the problems of normative thought—in this case, how the attitudes expressed by the premises are inconsistent with not having the attitude expressed by the conclusion.

But given this approach to the logical connective in **1**, we now have an easy decision to make: do we have one indicative conditional, or four? Compare the following:

- 1** If being friendly is wrong, then being friendly to strangers is wrong.
- 4** If being friendly is wrong, then Jo isn't friendly to strangers.
- 5** If strangers are dangerous, then being friendly to strangers is wrong.
- 6** If strangers are dangerous, then Jo isn't friendly to strangers.

Our expressivist treatment of **1** gave its meaning by telling us what mental state is expressed by the whole sentence as a function of the

---

<sup>19</sup> See especially van Roojen (1996), Unwin (1999 & 2001), Sinnott-Armstrong (2000), and Dreier (2006).

mental states expressed by the antecedent and consequent. So if *any* of the conditionals in **4**, **5**, and **6** are the *same* as the conditional in **1**, then it must be the *same* function from two mental states to a third mental state. And if so, then the non-normative sentences embedded in **4**, **5**, and **6** must *express* some mental state in the *same sense* as the normative sentences embedded in **1** do—*whatever* sense it is in which the logical connectives’ meaning is given by functions from the mental states expressed by embedded clauses to the mental state expressed by the sentence as a whole. So unless the expressivist genuinely wants to introduce four logical connectives for every logical connective in the surface grammar (or nine, or sixteen, or more, if there are more than two kinds of mental state “expressed” by sentences<sup>20</sup>), she is committed to holding that non-normative sentences express mental states in the *same sense* as do normative sentences. That is, she is committed to the *parity thesis*. The states expressed are simply beliefs, rather than non-cognitive attitudes.

## 2.2 Expressivist Mentalism

Given this, we can draw one more conclusion about the commitments of pure expressivism. Since **6** is wholly non-normative, we naturally expect that it has an ordinary, descriptive, propositional content, a proposition which is *about the world*. But since it gets its meaning *from* that of the conditional, and the meaning of the conditional is given by a function from the mental states expressed by the antecedent and consequent to the mental state expressed by the whole, it follows that it must get its propositional content *from* the mental state that it expresses.<sup>21</sup> But that’s okay. After all, the mental state expressed by **6** should turn out to be the belief that if strangers are dangerous then Jo isn’t friendly to strangers. And that *belief* has the propositional content that if strangers are dangerous then Jo isn’t friendly to strangers—that is precisely what distinguishes it from other beliefs. So on this view sentences like **6** must *get* their propositional contents *from* the mental

---

<sup>20</sup> There will be infinitely many, if the mental state expressed by a compound sentence such as **1** doesn’t fall neatly into the categories of belief or of disapproval (in the stipulative sense of “disapproval”), because conditionals can themselves be embedded in conditionals. This is true of both Simon Blackburn’s and Allan Gibbard’s considered accounts of which mental state is expressed by **1**, even though it was not true of Blackburn’s earlier accounts. See Blackburn (1988 & 1998), Gibbard (1990 & 2003), and compare Blackburn (1973 & 1984). So if Blackburn and Gibbard want a finite number of conditionals, then they need to assume that there is only one, which Gibbard clearly does.

<sup>21</sup> “[T]o explain the meaning of a term, explain what states of mind the term is used to express” Gibbard (2003, 6-7).

states they express. And this follows simply from the role that “if...then” plays in the language—that of assigning a *mental state* to sentences containing it, rather than simply assigning them a proposition.

So given the expressivist account of the logical connectives, all complex non-normative sentences get their propositional contents *from* the mental states that they express. So what, then, about *atomic* non-normative sentences? How do they get their meaning? Could it be that they get their propositional contents in some other way? Pending a lemma to be argued for in the next section, I think not. For we know that the belief that is expressed by a sentence has to have the same propositional content as that sentence. So it seems that there are three possibilities. Either 1) the sentence has the propositional content that it does because it expresses that belief, or 2) the sentence counts as expressing that belief because they have the same content, or 3) it will be a coincidence if the sentence has the same propositional content as the belief it expresses, because nothing guarantees it. But it is too much to believe that 3) could be true. Atomic non-normative sentences *have* to have the same propositional content as the beliefs that they express. And I will argue in section 3.1 that given the pure expressivist’s other commitments, 2) can’t be the case either. So given the advertised lemma to be discharged in section 3.1, it follows that *all* non-normative sentences with propositional contents inherit those contents from the beliefs that they express.

This view is a version of a roughly Lockean view that we might call *Mentalism*. Given that both beliefs and ordinary descriptive sentences both have propositional contents conveying information about the world, there are at least three broad shapes of view that we might take about how they get those contents. According to one extreme view, first language gets its content, and then people are able to have beliefs with propositional contents by accepting sentences with those propositional contents. According to the reverse extreme view, first beliefs get their contents, and then sentences come to have contents by being used to express beliefs. And on yet a third view, there are both sentences and beliefs, and what they all mean is determined jointly by facts about how people use them.

Mentalism is the second view. It is important to understand pure expressivism’s commitment to mentalism in order to appreciate the dissatisfaction that I will express with Gibbard’s account of the *expression* relation in part 4. But for now, given the promised lemma, we’ve uncovered a *lot* about the commitments of pure expressivism. Perhaps there is some attractive pure expressivist view that avoids these commitments. If so, I do not know what it is. But even so, it is enough for

me that the commitments I have described constitute an important and attractive option in metaethics. The questions I will be raising about the nature of the *expression* relation are questions about how *this* view is to deal with it. Whether or not it is the only version of pure expressivism, its attractions are clear enough that it is the only one that I will concern myself with in what follows.

The central characterization that we have been able to give of the principal attractions of pure expressivism is that it aspires to hold that there is nothing particularly distinctive or problematic about normative *language*, but only something distinctive and interesting about normative *thought*. After all, according to the parity thesis, the only distinctive thing about normative sentences is what *kind* of mental state they express, and we all know, presumably, what *expression* is—it is the relation that “grass is green” bears to the belief that grass is green. But what we will now see in what follows is that this attractive thought is, at the least, highly misleading. The cost of accepting the parity thesis for expressivism is high—in the philosophy of *language*. And in particular, it leads to a very surprising and highly controversial view about just what relation “grass is green” *does* bear to the belief that grass is green.

### 3

#### 3.1 *A first pass - Same Content Expression*

If the Big Idea of pure expressivism is really right that *whatever* explains why “grass is green” is not subject to modal and disagreement problems will work to explain why “murder is wrong” is not either, then presumably the most obvious, uncontroversial explanation of the difference between “grass is green” and “I believe that grass is green” should be amenable to expressivism. But as Philip Pettit and Frank Jackson have pointed out,<sup>22</sup> this is not the case. The most obvious story about how “grass is green” is related to the belief that grass is green is that both have the *same content*—that grass is green. But it is easy to see that this will not do for expressivism. Expressivism, after all, is trying to explicate the meaning of normative language *by appeal* to the non-cognitive role of normative thinking. But on this account of expression—the *same content* account—whether or not a sentence expresses a mental state needs to be explained by the propositional contents of both normative language *and* normative thought. So it gets things entirely the wrong way around.

---

<sup>22</sup> Jackson and Pettit (1997, 244-245), in responding to objections to their “problem” for expressivism. See section 5.1.

More carefully, we know that pure expressivism, at least as we are understanding it, is committed to the *parity thesis*, and to the view that at least *complex* non-normative sentences inherit their propositional contents from the mental states that they express. But together, these two views are transparently inconsistent with the *same content* account of *expression*. Mentalism about complex non-normative sentences precludes their expressing mental states in the *same content* sense because it requires that they express a mental state *in order to* have a propositional content, rather than conversely. And so from the parity thesis it then follows that the *same content* account cannot be right about how any other sentences express their respective mental states, either. And that is the lemma needed in my argument that expressivism is committed to holding that even atomic non-normative sentences inherit their propositional contents from the beliefs they express.

This is, already, a considerable blow to the Big Idea that *no matter what* explains why “grass is green” is not subject to modal and disagreement problems, it will work for “murder is wrong” as well. At the very least, it forces expressivists to take a view about what the relation *is* between “grass is green” and the belief that grass is green. So let’s keep looking for what that view should be.

### 3.2 *A Second Go—Causal Expressivism*

If we’re in the market for relations which “grass is green” uncontroversially does bear to the belief that grass is green, a natural place to begin looking is to the typical *causal* relationship between the belief that grass is green and at least paradigm utterances of “grass is green.” Uncontroversially, in normal cases, when the speaker is not lying, and is making a genuine assertion, there *is* some causal relationship between the belief and the utterance. And in contrast, normal utterances of “I believe that grass is green” are not caused in this way by the belief that grass is green. Except (possibly) in cases of so-called “guarded assertions” that grass is green, such utterances are caused by a belief *that one believes* that grass is green. So let us see how this does for the expressivist.<sup>23</sup>

Some have suggested to me that causal expressivism is not worth taking very seriously. But though I will explain shortly its defects, which I’ll suggest are irreparable, I do think that it is worth taking seriously. For one thing, expressivists have in general told us so little

---

<sup>23</sup> Obviously not just *any* way of being caused will do—an utterance must be caused *in the right way* by a belief in order to count as an expression of that belief. But for our purposes here, it doesn’t matter what “the right way” turns out to be. All such views will be subject to the same difficulties.

about what they mean by “express” that it is not safe ruling anything out, and Ayer, at least, was clearly a causal expressivist with respect to at least one of the four uses of moral language for which he allowed in *Language, Truth, and Logic*. Surely “ejaculations,” as he calls them, are *caused* by the emotions that they express.<sup>24</sup> Simon Blackburn also certainly suggests a causal reading when he characterizes expressivism as holding that when “assert values,” “we *voice* our states of mind.”<sup>25</sup>

For another, once we set aside the linguists’ stipulative technical use of “express”, in which sentence-types express propositions relative to contexts of utterance, common-sense intuitions about what attitude a *speaker* expresses by some activity, linguistic or otherwise, seem to commit to the view that you can only express attitudes that you have. It is natural to think that liars only *purport* to express their beliefs, expressing instead their deceitful intentions, and that whether a particular grimace is an expression of excruciating agony or sexual ecstasy depends on which its subject is undergoing at the moment. If expressivists really mean to be using “expression” in a sense that we all antecedently understand, there is ample evidence that we should take seriously the idea that what we antecedently understand is some causal relation.

There is yet one more reason to suspect that causal expressivism is worth taking seriously. Recall that the Frege-Geach problem was set up with the observation that I can assert **1** without expressing the attitude that I would express by asserting **2**:

- 1** If being friendly is wrong, then being friendly to strangers is wrong.
- 2** Being friendly is wrong.

But what makes it so obvious that I am *not* expressing the attitude that would be expressed by an assertion of **2** when I assert **1**? After all, an ordinary compositional truth-conditional semanticist will say that the proposition that being friendly is wrong *is* expressed in **1**—and that that is precisely why “if...then” can operate on that proposition in order to yield the proposition that if being friendly is wrong, then being friendly to strangers is wrong.

I take it that what makes it so obvious that I can assert **1** without expressing disapproval of being friendly, is that I can easily *have* the attitude expressed by **1** without *having* the attitude expressed by **2**. But that seems to be relevant only if there is some close connection between

---

<sup>24</sup> Ayer (1936, 103).

<sup>25</sup> Blackburn (1998, 50 (*italics in original*)).

expressing a mental state and being *in* it, as is claimed by the *causal* account of *expression*. So I think it is well worth taking causal expressivism seriously, in the sense of being very clear about the problems that it faces. And doing so will also help us to better understand why the later accounts are on the right track. I will list two such problems here. Both stem from the fact that causal expressivism makes it impossible to express a mental state that you are not in.

The first problem is that lies have the same propositional contents as their corresponding sincere assertions. It is precisely because they have the same contents that liars, by and large, manage to succeed at saying something *false*. Never mind whether there are normative lies. According to the parity thesis, utterances of non-normative sentences must express mental states in the same sense in which utterances of normative sentences do, and according to the expressivists' Mentalism, non-normative sentences get their propositional contents from the beliefs that they express. So from these views it follows that lies have the same propositional content as sincere utterances only if it is possible to express a mental state that you are not *in*. But according to the causal account of *expression* this is not possible—to express a mental state an utterance must be *caused* by it. So the causal account is wrong: fatal problem number one.

The second problem is that causal expressivism rules out the minimalist account of truth and falsity that appears to be necessary in order to make good on the expressivist solution to the hard disagreement problem. Recall that that was the problem of why negations of what another speaker has said, and claims that it is false, should be co-assertable. But “it is false that murder is wrong” is a sentence in which “murder is wrong” appears in an embedded context just as much as “murder is not wrong” is. So it is not hard to see how to complete the expressivist answer to this problem, by an extension of the pure expressivist answer we gave to the Frege-Geach problem. You do it by introducing the meaning of “false” by saying what mental state is expressed by sentences containing the word “false” as a function of the mental states expressed by the sentence to which falsity is being attributed. In order to solve the problem of disagreement, the idea is that the mental state expressed by “it is false that murder is wrong” should be one that a speaker is almost guaranteed to be in, if she is in the mental state expressed by “murder is not wrong.” All such views are *minimalist* treatments of “false,” and analogously of “true.” The simplest such view is the *redundant* view that the two mental states are the same.

But this minimalist account of “true” and “false” will not work for causal expressivism, and again because it rules out the possibility of

expressing mental states that you are not in. “True” and “false,” after all, are not only used in sentences like “the sentence ‘grass is green’ is true” and “it is true that grass is green.” As has been much-emphasized in the literature on minimalism about truth,<sup>26</sup> “true” and “false” are only particularly useful because they can figure in sentences in which the sentence to which truth or falsity is being ascribed is not spelled out—for example, as in “what Alexander said is true.” But a speaker can say that what Alexander said is true without even knowing what Alexander said, and hence while not holding any independent view about, for example, the GDP of Guatemala, if that is what Alexander turns out to have been talking about—or even while actively disagreeing. She just has to be confident that what Alexander said is true without knowing what it was that he said. And so the minimalist account works for these uses of “true” and “false” only if it is possible to express mental states that you are not actually in—indeed, to express mental states without knowing which mental states you are expressing. Second fatal problem for causal expressivism, and a moral: an adequate expressivist account of *expression* will have to make it possible to express mental states you are not in.

### 3.3 *Expression as Implicature*

Stephen Barker and David Copp have helpfully suggested that normative sentences might express non-cognitive attitudes by pragmatically conveying the *information* that the speaker is in that non-cognitive state as a matter of being a Gricean *conventional implicature*.<sup>27</sup> This view would have the advantage over causal expressivism that it makes it possible to express a mental state that you are not in. But though it may be a viable option for the adulterated expressivist,<sup>28</sup> it will not do for pure expressivism. For by the pure expressivist’s Mentalism, she holds that the mental state *expressed* by a sentence is no mere pragmatic overlay on language, but what explains how non-normative sentences come to have the propositional *semantic* contents that they do in the first place. Accepting this view would be like holding that “but” denotes conjunction *because* it is conventionally used contrastively, or that “and” denotes conjunction *because* there is a convention about temporal order matching sentential order in sentences like “she ran to the edge of the cliff and jumped.”

---

<sup>26</sup> Horwich (1990), Soames (1999), etc.

<sup>27</sup> Grice (1967).

<sup>28</sup> Barker and Copp are both what I have called adulterated expressivists. But see also Finlay (2005), who critically assesses their views.



Three up, three down. None of these obvious candidates for the *expression* relation will work for the pure expressivist. To find out what pure expressivists need to say about *expression* we need to look beyond the obvious.

## 4

### 4.1 Gibbard and Indicatory Expressivism

Allan Gibbard has a better idea about the way in which “grass is green” is related to the belief that grass is green. Granting his pure expressivist Mentalism, he knows that however sentences express mental states, non-normative sentences get their contents *from* the mental states that they express. And given this, obviously it must be possible for a sentence to express a mental state that its speaker is not in. So the *expression* relation must be in this sense *intentional*. And that makes Gibbard’s view a very natural one. On his view, it is a matter of the speaker’s *intentions*. Here is how he describes the picture in *Wise Choices, Apt Feelings*:

Here to be sincere is to express only beliefs that one actually has, and to be an authority on something is to be quite unlikely to be mistaken about it. Caesar thus intends to get Cleopatra to believe that he was captured by pirates in his youth, and to do so in the following manner. He utters words that conventionally purport to express, on the part of any speaker, a belief that he was captured by pirates in his youth. He intends her to come to accept that he has that belief, and to do so in virtue of her recognition of this intention. Since she takes him to be sincere, she has reason to accept, upon hearing his words, that he does believe he was captured by pirates in his youth. Since she thinks him an authority on his youth, she concludes from his believing it that he indeed was captured by pirates in his youth.<sup>29</sup>

Since on Gibbard’s view the mental state that an utterance expresses is the one that the speaker intends to indicate to his audience that he is in, I call it *indicatory* expressivism.

Since Gibbard is a pure expressivist, complete with all of its commitments, he holds that it is *by* indicating to an audience what mental state the speaker is in, that an utterance comes to have the ordinary propositional content that it has. And in the quoted passage, he explains how. By uttering the sentence, “I was captured by pirates in my youth,” Caesar tries to convey the information that he believes that he was captured by pirates in his youth. Since Cleopatra believes that

---

<sup>29</sup> Gibbard (1990, 85).

Caesar is being sincere, she infers that he really does believe this. And then, since she “thinks him an authority on his youth,” she concludes that the *content* of this belief is actually true—that Caesar was captured by pirates in his youth. And so this information—the propositional content of the sentence—is information that is conveyed *indirectly*, by means of conveying the information that Caesar is in a certain mental state.

This, obviously, is not an uncontroversial story about the relationship between “grass is green” and the belief that grass is green. It is very substantive story, but one that seems to be forced on expressivism by its commitment to Mentalism and to the parity thesis. By the standards of part 3 of this paper, it does fairly well. It respects the expressivist’s commitments to Mentalism and to the parity thesis, and allows for the possibility that an utterance can express a mental state that the speaker is not in—if the speaker is *insincere*. But we still might like to know more about this idea—that the mental state expressed by an utterance is the one that the speaker primarily intends to convey to his audience that he is in. If the mental state expressed is merely a matter of the speaker’s intentions—of what is going on in her head at the time—then it is natural to think that it could easily happen that a speaker happens to have these intentions with respect to sentences which do not, intuitively, have the appropriate semantic contents.

In *Thinking How to Live* Gibbard voices this question explicitly:

One further point about expressing: it might be thought that one could express a belief with an explicit performative, as with  
I hereby express the belief that Moriarty has arrived. (*E*)<sup>30</sup>

This is indeed what I might have thought, if what the speaker *expresses* is really just a matter of her *intentions*. For a speaker can certainly utter sentence (*E*) with the intention to convey to her listeners that she believes that Moriarty has arrived, hoping that since they think her an authority on Moriarty’s comings and goings, they will infer that Moriarty has arrived. I might even have thought that these sentences, and others, would do as well:

- 7 Allow me to convey to you the information that I believe that Moriarty has arrived.
- 8 That Moriarty has arrived is the content of a belief that I am using this sentence to inform you that I have.

---

<sup>30</sup> Gibbard (2003, 78), italics added for emphasis. Gibbard is here addressing a concern raised by van Roojen, 1996.

- 9 I'm hoping that by hereby letting you know that I believe that Moriarty has arrived you will infer that I really do believe it, and hence, that since I am an authority on Moriarty's comings and goings, he really has arrived.

Yet even if you did utter these sentences with such intentions, they would not *thereby* come to have, even on this one particular occasion, the semantic content that Moriarty has arrived.

Gibbard's answer to this *prima facie* objection in *Thinking How to Live* is not very satisfactory. Here is what he says immediately following:

But for the sense of the term 'express' that I intend, this isn't so. Holmes would normally express his belief that Moriarty has arrived by saying "Moriarty has arrived," and the logical force of this is quite different from the logical force of (E). For one thing, to pronounce (E) is to say something about what one is doing, whereas to say "Moriarty has arrived" is to do no such thing. But still, with (E) doesn't Holmes also express his belief that Moriarty has arrived? He may do something else as well, but doesn't he at least do this among other things? I think not. *If Holmes expresses a belief, then you contradict what he says by expressing a contradictory belief.* When he says "Moriarty has arrived," for instance, you contradict him if you say "Moriarty has not arrived." When Holmes utters (E), however, you don't contradict him if you say "Moriarty has not arrived." You contradict him by saying, "You don't thereby express that belief." In the sense of the term 'express' I intend, you express a belief by making a statement. You cannot, it now appears, do the same thing by uttering an explicit performative.<sup>31</sup>

This answer, I think, is clearly a step backwards. All of the things Gibbard cites here are evidence that on an *adequate* account of the *expression* relation, no utterance by an English speaker of Gibbard's sentence (E) will count as expressing the belief that Moriarty has just arrived. But it does not amount to any kind of *explanation* of what feature of Gibbard's own account of the *expression* relation rules this out.

The problem is that in answering this challenge, Gibbard is appealing to the things he needs to *explain*, in order to argue that his account must explain them. Now if Gibbard is right that there is an uncontroversial sense in which "that words express judgments will, of course, be accepted by almost everyone," then this is something to which he might rightly feel entitled to appeal. But the question that is now at stake is just what this uncontroversial sense is, in which words express judgments—that can be appealed to by "almost everyone," and will still be

---

<sup>31</sup> Gibbard (2003, 78-79).

consistent with the expressivist's commitments. In order to answer this question, we cannot simply point out that an account that failed on this score would not be an adequate account of the *expression* relation. We must show that there is an account of *expression* that works for the expressivist and *does not* fail on this score. And this is something that Gibbard does not do in *Thinking How To Live*.

On the other hand, this is not by itself a deep objection to Gibbard, because I think that in any case we can see how to answer this question on the basis of Gibbard's original remarks in *Wise Choices, Apt Feelings*. The idea is that examples like Gibbard's sentence (*E*) make it clear that not *just any* intentions of a speaker to convey information about her mental states to her audience will make the speaker count as expressing those mental states. We need some *restriction* on the speaker's intentions, and the restriction has to suffice to rule out all cases in which the sentence would end up having, intuitively, the wrong semantic content. But Gibbard's original account shows us how to do exactly that. For on Gibbard's original account, it is not just any intentions of the speaker that count, but only intentions to take advantage of *already existing linguistic conventions*. What Caesar does, says Gibbard, is "[h]e utters words that conventionally purport to express, on the part of any speaker, a belief that he was captured by pirates in his youth."<sup>32</sup>

I take it that some such qualification is necessary in order to fix indicatory expressivism to rule out sentences like Gibbard's (*E*) turning out to have the semantic content that Moriarty has arrived. According to this view, though what Gibbard cites in *Thinking How To Live* is *evidence* that the speaker has not expressed a belief that Moriarty has arrived, what *explains* why the speaker has not done so is that there are no conventions *of the right kind* in the language to indicate to your audience that you believe that Moriarty has arrived by saying, "I hereby express the belief that Moriarty has arrived" or the sentence "I hereby convey to you that I believe that Moriarty has arrived" or any other such sentence. The only sentence in English to which the *right kind* of such conventions are attached, is the sentence "Moriarty has arrived."

So indicatory expressivism seems to be committed, with cause, to the view that there are linguistic conventions which directly, all by themselves, convey information about speakers' mental states. And it is by exploiting these conventions and the fact that mental states, independently of language, have propositional contents about parts of the world other than the speaker's own mind, that speakers manage to

---

<sup>32</sup> Gibbard (1990, 85).

convey information about parts of the world other than their own minds.

#### 4.2 *An Unstable View*

The challenge I want to raise for indicatory expressivism is to this *package* of commitments, which are, I will suggest, uncomfortably unstable. According to the view, what languages do is set up a group of conventions in virtue of which speakers can convey information to each other by uttering sentences. And according to this view, speakers utter sentences with the intention of conveying the information that is conventionally attached to sentences, in this way. But according to this view, the information about the world that conventions assign to sentences so that it can be conveyed from speaker to listener by an utterance of the sentence is not the sentence's propositional content—sentences are not true or false when the information that is conventionally assigned to the sentence is true or false. Propositional content, according to the view, is an entirely different sort of thing. The information that is assigned to sentences by the conventions of the language so that it can be conveyed by speakers to listeners is not their *semantic* content, but something else—what we might call their *ur-content*. Propositional content is something that sentences get *in virtue* of their *ur-content*, when the *ur-content* is that the speaker has a belief with some propositional content.

But I think this view is unstable. The problem is that on this view, *ur-contents* turn out to be so *robust*, that it makes it hard to see why *ur-contents* aren't *normal* contents. Obviously *according to the expressivist* they are not normal contents, but we can get at the thought which bothers me by considering the following thought experiment. Suppose that there are a group of creatures without a language, but with the need to convey information to one another, and perhaps to do other sorts of thing—whatever it is that normative language allows groups to do. Imagine that they figure out how to establish conventions so that by uttering some string of phonemes, they can convey information to each other about the world. According to Gibbard's indicatory expressivism, all that they will have is *ur-contents*. But why isn't that enough? It lets them convey information about the world, which is what they set out to do.

Let's consider a couple of possible answers. Could the answer be that linguistic conventions are only rough-and-ready things, which can't, by themselves, discriminate between contents as fine-grainedly as can the contents of thought? That can't be. Because clearly, in order to discriminate between all possible mental states, *ur-contents* must be as fine-grained as propositional contents.

Another possible answer might be that there is some reason that linguistic conventions can only directly assign information about the speaker's own mental states, and so speakers need help from the contents of beliefs, in order to convey information about other parts of the world. But this does not seem to be a very plausible answer, either, for it is hard to see what this reason could possibly be. Try pointing at a stick and saying "stick," in order to establish a convention to use "stick" to refer to sticks. Then try establishing a convention to use a sentence to convey the information, for any context of utterance of which the speaker is  $x$ , that  $x$  believes that this is a stick. Which seems easier?

So it seems that because of its two main features, indicatory expressivism makes *ur*-contents robust enough that there should be a real puzzle about why we would ever have needed normal contents. Those two features are 1) that it makes *ur*-contents out to be propositions<sup>33</sup> that are conventionally assigned to sentences, and 2) that it holds that *ur*-contents are *information* that a speaker *conveys* to her audience by uttering the sentence. The problem is that indicatory expressivists don't *really* think that we need a distinction between *ur*-contents and normal contents in order to successfully convey information about the world with one another by means of linguistic conventions. They only think that we need this distinction, because it is necessary in order to allow normative and non-normative vocabulary to compose under the logical connectives, by giving non-normative sentences an *extra level* of content to correspond to the only kind of content that normative sentences have.

The instability of Gibbard's indicatory expressivism is that it attributes to *ur*-contents all that we might want of ordinary propositional contents, but at the same time denies that *ur*-contents have what it takes to be ordinary propositional contents. I don't pretend that this is anything that Gibbard doesn't know, or which would be sufficient to talk him out of indicatory expressivism. Nothing about the view commits Gibbard to collapsing the distinction between *ur*-contents and normal contents—it just leaves him with a puzzle about *why* we need such a distinction, if *ur*-contents are capable of doing everything that normal contents are designed to do. But I do think that it should strike us as colossally unsatisfactory, if there is any way at all that we could do better for pure expressivism. And so in part 5, I'll explain how I think we *can* do better.

---

<sup>33</sup> Strictly speaking, conventions assign sentences to something that is *character*-like, rather than propositional. Each sentence gets assigned by convention a function from contexts of utterance to *ur*-contents that are propositions about the mental states of the speaker of the context.

### 5.1 Assertability Expressivism

The problem, I think, is that Gibbard's indicatory expressivism commits to more than is necessary in order to get out of the problems raised for the accounts of *expression* in part 3, and in two ways. The main moral of that section, recall, was that the *expression* relation needs to be *intentional*, in the sense of leaving it possible to express a mental state that you are not in. And Gibbard proposed that it is a matter of the speaker's *intentions*. But that, I think, is overcommitting. Moreover, even exploiting the proposal shared by Gibbard and the implicature expressivists, that a sentence expresses a mental state by being associated with a *proposition* to the effect that the speaker is in that mental state, both Gibbard and the implicature expressivists go further than necessary, by suggesting that the proposition so associated with the sentence has to be one that is *conveyed* to listeners, whether intentionally or not.

So *assertability* expressivism aspires to give a *minimal* account of the *expression* relation that avoids making these two mistakes. It shares with Gibbard and with the implicature expressivists the idea that sentences are associated with propositions—characters, really—to the effect that the speaker is in some mental state or other. According to assertability expressivism, the mental state expressed by a sentence is the one that gets mentioned in this proposition—its *ur*-content. But assertability expressivism denies not only that *ur*-contents need be something that the speaker utters a sentence with the intention of conveying to her listener, it denies that they need be information that is conveyed *at all*. According to assertability expressivism, they are just the *assertability conditions* of the sentence, a device of the semantic theorist.

The idea is this. Every sentence in the language is associated with conditions in which it is *semantically correct* to use that sentence assertorically. But we are not to think of those conditions as *truth* conditions so much as *use* conditions. After all, a speaker who is mistaken about who is president of the United States, believing it to be John McCain, is not making a *linguistic* mistake by asserting, “the president of the United States is John McCain.” She is making a *mistake*, but not a *linguistic* one, since she really does believe that John McCain is president. So semantic correctness conditions capture the conditions under which a speaker would be making a *linguistic*, *semantic* mistake

in the use of some sentence, which are not the same as its truth conditions.<sup>34</sup>

Moreover, according to assertability expressivism, since it accepts the expressivist's Mentalism, these assertability conditions are not assigned to sentences *derivatively*, by means of their truth conditions. Rather, sentences *get* their truth conditions *from* their assertability conditions, by inheriting them from the truth conditions of the belief mentioned in the assertability conditions.

Why is it, then, that assertability conditions must always mention a mental state of the speaker? Why can't they be about other parts of the world? To that I can give two different kinds of answer on behalf of assertability expressivism. The first is to appeal to the so-called *norm of assertion*. It is periodically discussed what the norm of assertion is, in the sense of when it is correct to assert some sentence.<sup>35</sup> Possible answers include: when you believe it, when you justifiably believe it, when you know it, etc. The idea presupposed by this literature is that sentences have a certain propositional content, and then there is a background rule about assertion which says, for example, to assert that sentence only if you have the belief with its propositional content. But assertability expressivism explains the norm of assertion the other way around. According to assertability expressivism there is no single rule governing assertion *as such*. Rather, each sentence comes with rules about when *it* may be asserted. And since these mention when the speaker comes to have some belief, the sentence comes to count as having the content of that belief derivatively. So it is still true that for any sentence with the content that *p*, it is correct to assert that sentence only if you believe (justifiably believe, know, etc.) that *p*. But instead of this being true because the sentence means that *p* and there is a general norm governing assertion, we simply say that the sentence means that *p* because that is the belief that it is only correct to assert the sentence if you have.

---

<sup>34</sup> It might be objected that this can't be quite right, because even if lies violate "assertability conditions," they do not seem to be *semantically* incorrect. The assertability expressivist can offer one of two responses. She can say that the talk about "semantic correctness" is not really to the point—what is important is only that there are assertability conditions for the sentence. Or she could continue to maintain that assertability conditions *are* semantic correctness conditions, and accept that lies *do* violate semantic correctness conditions—the rules of the language. Plausibly, liars *do* break the rules of the language: they do so *deliberately*. Situations in which lying is appropriate, such as in the game called "Bullshit," might either be treated as cases in which other considerations make it permissible to break the rules of the language, or, plausibly, as cases of less than full-blooded assertion. Thanks to an anonymous referee for raising this point.

<sup>35</sup> See, for example, Williamson (2000, 238-269).



This, I think, is a very indirect argument that assertability conditions will always mention mental states. It doesn't *explain* why this is true, but uses this idea to explain natural intuitions that there is a *norm of assertion*. But I think we can do better, by appeal to the idea that assertability conditions give the rules under which speakers will count as making *linguistic mistakes*. If assertability conditions are not to fail due to an error of the speaker's about something other than language, then they will have to be something to which speakers have a special kind of *access*. And so it makes sense to suppose that they will be conditions of the speaker's own mind. And that explains *why* assertability conditions always mention a mental state.<sup>36</sup>

Assertability conditions, so conceived, are a device of the semantic theorist. They are not a kind of *information* that speakers intend to *convey*. So there is no sense in which a community of speakers could get by, managing to communicate information to each other about the world, by means of assertability conditions alone. It is only because some assertability conditions mention beliefs, and beliefs have contents about the world, that speakers can manage to convey information about the world. And that is how assertability expressivism gets around the unsatisfying instability in Gibbard's indicatory expressivism.

What we know about assertability expressivism is now enough to answer a challenge to expressivism raised by Frank Jackson and Philip Pettit<sup>37</sup>, in an article that has generated a good deal of attention. Jackson and Pettit claim to offer an argument that expressivism collapses into what I have been calling Cognitivist Speaker Subjectivism. In a nutshell, their argument is that language must be rule-governed, and so even normative language must be governed by the rule to assert a normative sentence just in case you are in the proper non-cognitive state. But similarly, they suggest, non-normative language is rule-governed in the following way: the rule governing "grass is green" is to assert it when grass is green. So, they argue, since given this rule "grass is green" has the propositional content that grass is green, by analogy it follows that "murder is wrong" has the propositional content that  $x$  disapproves of murder, where  $x$  is the speaker of the context. And so, they claim, expressivism collapses into CSS.

But it is now easy to say what is wrong with Jackson and Pettit's argument. Assertability expressivists agree that language must be rule-governed. But they hold that Jackson and Pettit have told the wrong

---

<sup>36</sup> The assertability expressivist's view about the foundations of semantic content is not unrelated to the views of Dummett (1975), (1976), (1977), to whom aspiring assertability expressivists might look for more help than I've been able to give, here.

<sup>37</sup> Jackson and Pettit (1998).

story about the way that “grass is green” is rule-governed. According to the pure expressivist, the rule governing “grass is green” is not to assert it when grass is green, but to assert it when you *believe* that grass is green. Since that doesn’t make the sentence turn out to mean that the speaker believes that grass is green, the rules governing the use of “murder is wrong” don’t make it turn out to mean that the speaker disapproves of murder. Now Jackson and Pettit may disagree with the assertability expressivist about this. They may hold that an utterance is semantically correct—in accordance with the rules of the language—just in case it is true, rather than just in case the speaker has the right belief. But that is just assuming something that we’ve now shown in detail that the expressivist has good grounds to reject.

Granted, this view about ordinary, non-normative language is highly controversial, and may certainly be false. But it’s question-begging to simply argue against expressivism by *assuming* that it is false. Assertability expressivists have their own coherent, and equally Lockean view about the rules governing language.<sup>38</sup> They simply hold that it is a mistake about the presidency, not about the words “president” or “John McCain”, if someone who believes that John McCain is president says, “John McCain is president,” and consequently that linguistic rules track, at least in the *first* instance, the attitudes of the speaker, and only derivatively who is president.<sup>39</sup>

## 5.2 *Where We Are*

In this paper I’ve tried to introduce the question as to what pure expressivists should think about *expression*. It is an interesting and important question both because expressivists have in general said so little about it, mostly taking it for granted that we already know, and because it plays such a central role in the attractions of expressivism in the first place. I’ve argued for a couple of key commitments to which we should understand pure expressivism as being committed, and tried to motivate the idea that pure expressivism should be understood as largely motivated by the Big Idea that normative language will not raise any new or deep problems not raised by normative *thought*. And the moral of the paper is that this Big Idea is, at the least, highly misleading. Expressivists buy the parallelism in their treatments of normative and non-normative language not by modeling normative language

---

<sup>38</sup> It is Lockean in its Mentalism. But of course expressivists accept Mentalism at the level of complete sentences, rather than at the level of individual words/concepts, as Locke did.

<sup>39</sup> See Dreier (2004) for the best available discussion of Jackson and Pettit’s argument.

on uncontroversial features of non-normative language, but by adjusting their account of non-normative language in order to make it fit seamlessly with their account of normative language. And this is a particularly high cost to pay in the philosophy of *language*.

If pure expressivism is going to get off of the ground, it cannot do with any of the uncontroversial relationships between “grass is green” and the belief that grass is green. It needs to marshal its creative resources in order to construct a highly controversial theory about how “grass is green” might be related to the belief that grass is green. And if what I’ve been arguing here is correct, then the most promising way for expressivists to do that is precisely the minimal one, in order to achieve the required results. It is to adopt *assertability* expressivism. I haven’t been out, here, to object to assertability expressivism. In fact, I’ve been careful to respond to some of the more obvious objections, and to set it up as persuasively as possible. But it should be clear that it is a highly controversial view about the semantic content of ordinary, non-normative language. It is not at all clear that “almost everyone” will agree that words express judgments in *this* sense.

The picture that I’m pushing is that since pure expressivism is only as palatable as its commitments in the philosophy of language, we should *use* those commitments in order to evaluate the view. I don’t claim to have done that in this paper, but only to have taken a pass at a prerequisite: arguing for *what* those commitments *are* in as transparent a way as possible, in order to make clear what the issues are. If I am right, then pure expressivism should be thought of as assertability expressivism. Is that good? Bad? Outrageous? I leave it for others to say, or myself on a later occasion. But I’ll close by raising a couple of the new issues that I think assertability expressivism forces on us.

### 5.3 *New Issues*

First, according to assertability expressivism to say that a normative utterance expresses a non-cognitive state of mind is to say that it is *semantically correct* to assert that sentence only if you are in that state of mind. So this raises the obvious question of whether *semantic correctness* is a normative concept. If it is, then assertability expressivism falls under its own scope, and the central claims of the theory turn out to be mere (?) expressions of non-cognitive states of mind. Is that an objection? I can’t decide.<sup>40</sup> But if what I’ve argued is right, then there is

---

<sup>40</sup> Jamie Dreier’s (2002) discussion of a related problem suggests that he thinks it would be, but he offers a way of getting out of the difficulty for the expressivist.

significant pressure deriving from a very general source pushing expressivists in this direction.<sup>41</sup>

Second, it is periodically argued that despite the protestations of expressivists, expressivism is still really committed to the view that morality is in some way “relative,” depending *in some sense or other* on the attitudes of speakers.<sup>42</sup> We might try, for example, to revive the modal and disagreement objections by appeal not to the notion of *truth* but that of *semantic correctness*. So now you speak truly when you say, “if I didn’t disapprove of murder, then it would be semantically incorrect for me to say that murder is wrong.” And though Sally can disagree with Jim by saying that what he says is false, she must still agree that it is semantically correct. Does that still sound bad? Maybe. Given the sense of “semantic correctness” that I tried to spell out in the last section, I’m not sure that it does. But if it does, does assertability expressivism have the resources to respond?

At first, I thought that it would. The idea would be to transform assertability expressivism from a view that validated the intuitive idea that belief is the norm of assertion, to one that validated the idea that *knowledge* is the norm of assertion. The idea would be that you would only have to admit that your counterfactual self spoke in a way that was semantically correct, if you were willing to ascribe *truth* to what your counterfactual self said. But given the minimalist treatment of “true” and “false,” you would *not* then allow that your counterfactual self spoke in a way that was semantically correct. At first, I thought that this idea would work, and yield a rich and resourceful brand of expressivism, if only it could make good on the generalized sense of “knowledge” to apply to the normative case. But now I wonder whether this will leave the expressivist unable to formulate her own theory—it was important, after all, to distinguish between assertability conditions and truth conditions when formulating assertability expressivism, but this solution requires insisting that semantic correctness conditions *entail* truth conditions. But I’m uncertain about the terrain, here; it requires more exploration.

I don’t pretend to have so much as raised an objection, here—at best, to have pointed at some things that need to be thoroughly investigated if we are to evaluate the prospects for assertability expressivism, and hence, if I am right, if we are to evaluate the prospects for pure expressivism full stop. But that is all that I set out to do. That words express judgments, I hope to have shown, is not something that we can all agree on, but the central controversial commitment of expressivism.

---

<sup>41</sup> Compare Rosen (1998).

<sup>42</sup> Zangwill (1994), Shafer-Landau (2003, 30-33).

This paper is just one step in figuring out what that means, and to what it commits expressivists.<sup>43</sup>

### References

- Alm, David (2000). "Moral Conditionals, Non-Cognitivism, and Meaning." *The Southern Journal of Philosophy* 38(3): 355–377.
- Ayer, A.J. (1936). *Language, Truth, and Logic*. New York: Dover.
- Barker, Stephen (2002). "Is Value Content a Component of Conventional Implicature?" *Analysis* 60(3): 268–279.
- Blackburn, Simon (1973). "Moral Realism." Reprinted in Blackburn (1993), 111–129.
- (1984). *Spreading the Word*. Oxford: Oxford University Press.
- (1988). "Attitudes and Contents." Reprinted in Blackburn [1993], 182–197.
- (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- (1998). *Ruling Passions*. Oxford: Oxford University Press.
- Cappelen, Herman, and Ernest Lepore (2005). *Insensitive Semantics*. Oxford: Basil Blackwell.
- Copp, David (2001). "Realist-Expressivism: A Neglected Option for Moral Realism". *Social Philosophy and Policy* 18(1): 1–43.
- Dreier, James (2002). "The Expressivist Circle: Invoking Norms in the Explanation of Normative Judgment". *Philosophy and Phenomenological Research* 65(1): 136–143.
- (2004). "Lockean and Logical Truth Conditions". *Analysis* 64(1): 84–91.
- (2006). "Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution." *Oxford Studies in Metaethics* 1: 217–233.
- Dummett, Michael (1975). "What is a Theory of Meaning? (I)." Reprinted in Dummett (1993), 1–33.
- (1976). "What is a Theory of Meaning? (II)." Reprinted in Dummett [1993], 34–93.
- (1977). *Elements of Intuitionism*. Oxford: Oxford University Press.
- (1993). *The Seas of Language*. Oxford: Oxford University Press.
- Finlay, Stephen (2005). "Value and Implicature." *Philosophers' Imprint* 5(4), <http://www.philosophersimprint.org/005004/>.
- Frege, Gottlob (1950). *The Foundations of Arithmetic*, trans. J.L. Austin. Evanston, IL: Northwestern University Press.
- Geach, P.T. (1960). "Ascriptivism". *Philosophical Review* 69: 221–225.
- (1965). "Assertion". *Philosophical Review* 74: 449–465.

---

<sup>43</sup> Special thanks to Mike McGlone, Jeff Speaks, and Amy Challen, and to the anonymous referees for *Philosophy and Phenomenological Research*.

- Gibbard, Allan (1990). *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.
- (2003). *Thinking How To Live*. Cambridge: Harvard University Press.
- Grice, H.P. (1967). “*Logic and Conversation*.” Reprinted in his *Studies in the Way of Words*. Cambridge: Harvard University Press, 1989, 1–144.
- Horgan, Terence and Mark Timmons (2006). “Expressivism, Yes! Relativism, No!” *Oxford Studies in Metaethics* 1: 73–98.
- Horwich, Paul (1990). *Truth*. Oxford: Oxford University Press.
- Jackson, Frank, and Philip Pettit (1998). “A Problem for Expressivism”. *Analysis* 58(4): 239–251.
- Joyce, Richard (2002). “Expressivism and Motivation Internalism.” *Analysis* 62(4): 336–344.
- Kaplan, David (1977). “Demonstratives.” In J. Almog, J. Perry and H. Wettstein, eds., *Themes from Kaplan*. Oxford: Oxford University Press, 1989, 481–563.
- Miller, Alexander (2003). *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press.
- Ridge, Michael (2006). “Ecumenical Expressivism: Finessing Frege.” *Ethics* 116(2): 302–336.
- Rosen, Gideon (1998). “Blackburn’s Essays in Quasi-Realism”. *Nous* 32(3): 386–405.
- Schroeder, Mark (ms). “Finagling Frege, Gaffing Geach.” Unpublished paper.
- Searle, John (1962). “Meaning and Speech Acts”. *Philosophical Review* 71: 423–432.
- (1969). *Speech Acts*. Cambridge: Cambridge University Press.
- Shafer-Landau, Russ (2003). *Moral Realism: A Defence*. Oxford: Oxford University Press.
- Sinnott-Armstrong, Walter (2000). “Expressivism and Embedding”. *Philosophy and Phenomenological Research* 61(3): 677–693.
- Soames, Scott (1999). *Understanding Truth*. Oxford: Oxford University Press.
- Stevenson, C.L. (1937). “The Emotive Meaning of Ethical Terms.” Reprinted in Stevenson (1963), 10–31.
- (1944). *Ethics and Language*. Oxford: Oxford University Press.
- (1963). *Facts and Values*. Westport, CT: Greenwood Press.
- Unwin, Nicholas (1999). “Quasi-Realism, Negation, and the Frege-Geach Problem”. *Philosophical Quarterly* 49(196): 337–352.
- (2001). “Norms and Negation: A Problem for Gibbard’s Logic.” *Philosophical Quarterly* 51(202): 60–75.

- van Roojen, Mark (1996). "Expressivism and Irrationality". *The Philosophical Review* 105: 311–335.
- (2005). "Expressivism, Supervenience, and Logic". *Ratio (new series)* 18(2): 190–205.
- Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.
- Zangwill, Nick (1994). "Moral Mind-Independence". *Australasian Journal of Philosophy* 72(2): 205–218.