The Frege-Geach Problem

Jack Woods

Non-cognitivist views of normative discourse, especially *emotivism*, *prescriptivism*, and their mutual successor *expressivism*, face a number of challenges in accounting for normative thought and talk. Perhaps the most prominent of these, "the Frege-Geach problem", is the challenge of explaining the content of normative thought and talk in complex constructions when the content of simple normative thought and talk are given non-cognitive treatment.[1]

The Frege-Geach problem arises for any view that takes force or analogous notions such as *expression* to explain content. For expressivists, the problem manifests in the fact that even if "Murder is wrong" expresses a conative state like disapproval in simple contexts, it does not do so in complex contexts like "If murder is wrong, it's not profitable." So the content of "murder is wrong" cannot be what is expressed. Explaining how the content is related to what is expressed—solving the Frege-Geach problem—is mandatory for a satisfying expressivist theory; the inability to so explain would be a damning objection to the view.

Following Frege, let us take the *content* of "the cat is on Matt" to be a certain picture of the world—one in which the relevant cat is on the relevant Matt—and the *force* to be my putting forth that the world is this way. I might utter the same *sentence* with different force and in certain complex descriptive constructions, the force of the component is *cancelled* entirely. "If the cat is on Matt, he'll get hives" does not describe the world as being such that the cat is on Matt, but rather describes a connection between one way the world could be and another. Frege

---

[1] I use "content" instead of "meaning" throughout to stress there is not a shared conception of linguistic meaning neutral between the views discussed.

used "⊢", a combination of a vertical "judgment stroke" and horizontal "content" stroke to indicate a forced expression. $\vdash p$ is a judgment with the content $- p$. Embedding, we obtain something like

$$\vdash \text{if} - p, \text{then} - q$$

where the indicative force, indicated by |, "scopes" out to the entire conditional, whose content is a connection between two ways the world could be, and which, when asserted, describes the world as obeying this connection. But the descriptive content of $- p$ is the same in the conditional just mentioned, in a simple assertion like $\vdash p$, and even in an interrogative utterance of p. So the force of an utterance separates from its content.

Frege argued on this basis that uttering "the cat is not on Matt" was not a case of *denial*, but an assertion in its own right, albeit an assertion of negated content. If we understand uttering this sentence as an act of *denying* that the cat is on Matt—paralleling the act of asserting—then we owe an account of what happens when we embed this sentence into a conditional like "if the cat is not on Matt, then he won't get hives". This conditional is asserted, not denied, and we neither assert nor deny the antecedent. For an assertion $\vdash p$, we can strip the force off when we embed it, leaving $- p$. If we try to do this with "the cat is not on Matt", understood as an act of denial, we end up with "the cat is on Matt" and this is plainly incorrect. So we cannot understand "the cat is not on Matt" in the conditional as an act of denial; when embedded, it simply does not have this force.

The problem reappears in Geach's "Ascriptivism" where he argued that R.M. Hare's *prescriptivist* account of moral discourse fails for similar reasons (Geach 1960). Hare analyzed the content of moral utterances largely in terms of their *prescriptive* force—their use in prescribing or permitting actions. Contemporary expressivists similarly use the role of moral utterances in giving voice to our evaluative or conative attitudes as the locus of their content. In force-canceling contexts like the antecedents of conditionals, however, the content of moral sentences is different from what it is in simple contexts. Embedded uses of moral sentences do not prescribe or express. This problem creeps as well; things we expect content to explain, like the validity of arguments, are also affected. Consider:

(P1) Murder is wrong.

(P2) If murder is wrong, then stealing is permissible.

(C) So, stealing is permissible.

If the content of 'murder is wrong' in P1 and in P2 differ, then we seemingly cannot compose P1 and P2 in order to derive C. To do so would, *prima facie*, be analogous to inferring that Barclays is the shore of a body of water from the fact that Barclays is a bank and banks are the shores of bodies of water.

Many theorists have taken the Frege-Geach problem to be the task of specifying a relationship of *validity* or *inference* which applies to statements when their content is understood in terms of force or expression. This potentially conflates a criterion for a successful solution with a solution itself. We need to specify an adequate account of content, in terms of simple force-involving assertions or judgments, with is adequate for expressions embedded both in complex constructions like the antecedants of conditionals and alternatively forced contexts like questions. Any reasonable account will give rise to a notion of validity which legitimates the intuitive relationships between logically complex normative sentences. The action of inference, however it is to be understood, should partially derive from the account of validity. Specifying a notion of validity directly thus seems the wrong way to approach the problem—unless logical inferentialism is correct—since it wouldn't explain *why* certain combinations of normative statements were inconsistent.

Initial reason to hope that we can solve the problem arises from the fact that *imperatives* can be embedded (Hare 1952). Even though 'if there's beer, get me one' doesn't itself express a command, the meaning of 'get me a beer' is clearly exhausted—or nearly exhausted anyways—by its use in expressing a command to the relevant party to get me a beer. And, whatever explains this would presumably also explain embedding normative content. But hope dwindles on inspection—constructions like 'murder is wrong' cleanly embed where imperatives do not. Compare:

1c If there's beer in the fridge, get me one.

2c If murder is fun, it's wrong.

1a #If get me a beer, there's beer.

2a If murder is wrong, it's fun.

3

We have no problem understanding the consequent embeddings (1-2c), but the antecedent embedding (1a) is incoherent, in contrast to (2a). So the analogy to imperatives will not get us very far. We need to actually develop an account instead of presuming that we can simply poach whichever explains embeddings in imperatives.

We survey some such answers to the problem shortly, but first we will discuss criteria for a successful solution. We focus on the Frege-Geach problem for expressivism about moral or normative thought and talk, putting other contexts to the side. The majority of what we say easily transfers.

Criteria for a solution to the Frege-Geach problem

Explaining Content

Explaining content in terms of what is expressed does not mean identifying content with what is expressed. Consider, as a similar example example, the Gricean explanation of descriptive meaning in terms of communication of our intentions of getting our interlocutor to believe something. My intention to get you to believe that the cat is on Matt is not itself the meaning of "the cat is on Matt", even though it plays an essential explanatory role in explaining the meaning. Rather, the meaning is the actual content of the belief we intend to get our interlocutor to adopt (Grice 1957).

Similarly to the Gricean, the expressivist attempts an account of content for some fragment of our thought and talk which satisfies the constraint that expressive use is fundamental to understanding such content in any context, simple or complex. They can satisfy this constraint by explaining the meaning of "murder is wrong" in embedded contexts by appeal to what it expresses in simple contexts without identifying the two (see (Hare 1971, p. 93), (Searle 1969), and (Sinclair 2011) for useful discussion).

Not all "content" need be expressible

An account of the content of embedded expressions, say in terms of states of mind, need not be reversible in the sense of providing an expression which corresponds to every state of mind. We need an *embedding* of normative language into states of mind, not a one-to-one mapping. This holds quite generally: consider standard

possible worlds semantics where we identify *propositions*—taken as the content of interpreted sentences—with sets of worlds. Plausibly, there are least $\aleph_0$ many worlds, so there are at least $2^{\aleph_0}$ many sets of worlds and correspondingly, at least $2^{\aleph_0}$ distinct propositions. But there are only $\aleph_0$ many sentences and, of course, $\aleph_0 < 2^{\aleph_0}$. So there can be no one-to-one mapping of sentences into propositions, so understood.

Likewise if we map normative sentences into something like commitments to possessing certain combinations of attitudes. We need not show that any particular combination of attitudes we can be committed to holding is expressible by a normative sentence of natural language. This mistake, most recently made in (Skorupski 2012), conflates the project of explaining normative content in terms of expression with the dubious project of *identifying* normative content with what is expressed. These should not be conflated—expressivists and fellow travelers use reasonably well-understood social, communicative, and psychological structure to explain the content of normative language and thought. They are not attempting to identify normative language and thought with this structure.

Mixed Expressions

Solving the Frege-Geach problem doesn't just mean giving an account of complex *normative* expressions—stealing is impermissible and murder is *really* impermissible—but also accommodating cases where we have composites of both normative and non-normative expressions—stealing is impermissible, but *really* fun. For example, it is not at all obvious what is expressed by assertions of disjunctions of moral and descriptive expressions: conative states, beliefs, both?

There are obvious hurdles to the tempting route of treating normative and non-normative expressions differently. Suppose, for example, we treat moral conditionals one way and descriptive conditionals another. Consider, then, the blatantly moral conditional 'if stealing is fun, it's not permitted' and its clearly descriptive contrapositive 'if stealing is permitted, it's no fun'. So ordinary logical transformations such as contraposition can change topic and, thereby, change what is expressed, even though we might want the two conditionals to be meaning equivalent (this depends on taking contraposition as valid for indicative conditionals.) On the other hand, treating normative and non-normative expressions alike may mean accepting non-cognitivism about descriptive expressions like "the cat is on Matt" which seems tantamount to a reductio of the approach (noted by Schroeder

(2008) before he heroically attempts to provide such an account.) Telling *some* story which evades these worries is mandatory for any resolution of the Frege-Geach problem.

Summary of Criteria

Summing up, the Frege-Geach problem is part of the development of an adequate non-cognitivist treatment of moral thought and talk. It is the task of giving an account of the content of expressions which:

- Assigns a content to both embedded and unembedded occurrences of expressions, subject to the proviso that the embedded occurrences are appropriately related to the expressive or forced unembedded occurrences.

- legitimates the logical relations between the expressions so analyzed.

- accommodates intuitive data about legitimate and illegitimate embeddings (as in the case of imperatives).

- Can accommodate mixed cases such as 'murder is wrong and fun'.

Furthermore, though it is not a no-go requirement, it is desirable if the resulting account is *compositional* in the sense that the content of complex expressions is a function of the content of the parts, *finitely specifiable* in the sense that we can lay down a set of rules for generating complex expressions on the basis of the parts, and *general* in the sense that we need not lay down special conditions to deal with complexities arising from each additional manner of composition accommodated. We now turn to describing the most prominent extant solutions. We start with minimalist solutions, turn to *discordance solutions* which posit some type of conflict between expressed mental states, then close by evaluating the resulting landscape.

The Metasemantic Gambit

Some theorists, such as Charlow (2013), Chrisman (2015), and Ridge (2014), argue that expressivism should be understood *metasemantically* (as a theory of how normative thought and talk has the content it has). Others, such as Schroeder (2008) argue that expressivism is best interpreted as a *semantic* view (a theory of normative content). Metasemantic expressivists typically claim that we can adopt a more-or-less standard semantic theory in accounting for the meaning of complex

normative utterances. This might be truth-conditional (Ridge) or more closely modeled on how we might assign meanings to imperatives (Charlow). Typically, however, these views solve the *semantic* Frege-Geach problem by showing that there is a natural compositional account of the meaning of complex expressions given the chosen semantic theory (Schroeder, though, argues that this undermines the use of semantic value in explaining normative judgment).

We might, however, worry that this just shifts the bump in the rug. After all, formal semantics need to be interpreted and interpreting aspects of a formal semantics can be rather difficult. See (Alwood 2015), for example, for useful discussion of how Ridge's view fares on this score. Dummett once complained that Davidson's use of a compositional truth-theory as a theory of meaning did not yet, in any substantive sense, provide an adequate account of linguistic understanding (Dummett 1975). Expressivists face a similar problem in providing a metasemantic theory which simultaneously legitimates the use of a standard compositional semantic theory (solving the Frege-Geach problem) while remaining both independently plausible and true to the spirit of the expressivist program. Some metasemantic expressivists, such as Ridge (2014), have a relatively easy time with the Frege-Geach problem, but this is more due to the hybrid aspect of their views (on which, see below) than the metasemantic aspect.

Expressivists have been less than fully clear about whether their program is semantic or metasemantic and less than fully clear about what they take linguistic meaning to be. A quick look at any traditional expressivist theory reveals that expressivists are typically more interested in metasemantic issues than giving a down-and-dirty formal semantics; accordingly, many expressivist views can be easily augmented with a deflationary theory of truth to provide a more-or-less standard truth-conditional formal semantics or are already cast halfway between a semantic and a metasemantic approach. On any expressivist view, we need to be able to provide an account of content of complex expressions; this might mean providing a compositional theory of mental states, a way of interpreting the semantics of complex expressions, or another direct account of the meaning of complex normative expressions. As we will see below, we face some version of the Frege-Geach problem on any way of proceeding.

Extant Solutions to the Frege-Geach Problem

Minimalist and Inferentialist Solutions

Hare mooted an early version of the inferentialist strategy to handle conditionals:

> To understand the "If... then" form of sentence is to understand the place that it has in logic (to understand its logical properties). (Hare 1971)

Hare's idea is that all that is needed to grasp the meaning of the conditional is to recognize the acceptability of inferring the consequent from the pair of the conditional and antecedent (or, perhaps, the disposition to do so.) Similar strategies could be used to explain most logical operations, drawing on existing work in the logical inferentialist tradition. If that's all that's needed to understand conditionals and other complex constructions embedding normative language, the Frege-Geach problem disappears. That is, it disappears as long as inferentialism is a viable program for explaining operators like the conditional.

A related strategy employs *minimalism* about truth in order to explain the content of complex constructions (Stoljar 1993). If we are willing to account for the truth of a statement ϕ in terms of the disquotational schema:

$$\phi \text{ if and only if } `\phi\text{' is true}$$

then we can apply ordinary compositional truth-laws in analyzing logically complex constructions like 'if the cat is on Matt, murder is wrong'. We analyze a conditional like the proceeding as true just in case the consequent is true if the antecedent is and treat "murder is wrong' is true' as equivalent to 'murder is wrong'.

Unfortunately, these solutions face numerous difficulties. Neither inferential-role nor a minimalist analyses will, by themselves, yield a plausible account of the meaning of normative language, expressivistically understood, in contexts like the antecedents of conditionals. We can bring this out by means of an example adapted from (Dreier 1994). Let 'Bob is hiyo' be a sentence whose primary function is to greet Bob. That is, when we see Bob and wish to greet him, we shout 'Bob is hiyo'. Now consider:

If Bob is hiyo, then he's in Princeton

Even if we allow that "Bob is hiyo' is true' is meaning-equivalent to 'Bob is hiyo', saying that the quoted sentence is true just in case Bob is in Princeton if he is hiyo—i.e. applying the minimalist schema to the components of the conditional—doesn't seem to explain it. Likewise, we do not have a clear sense of how and why someone would infer that Bob is in Princeton from the speech act of greeting Bob. These explanations seem incomplete—when we stipulate an expression with superficially descriptive grammar which is explicitly understood in terms of its force or expressive role, we have difficulties in understanding what it would mean to embed it in complex contexts *even given* a deflationist account of truth or a settled inferential role. So minimalism doesn't suffice to explain the content of *the components* in embedding contexts, even if we grant that it specifies a truth-condition for the complex construction, since the analysis of the truth of normative components like 'murder is wrong' outside of embedding contexts leans on what asserting that murder is wrong expresses. But in the embedded context, we have not expressed anything.

Discordance solutions

We start with a distinction, made vivid in (Schroeder 2008), between A-type and B-type solutions to the Frege-Geach problem. The issue concerns the way in which a set of attitudes can be inconsistent, or, more precisely, *discordant* (Baker and Woods 2015). Since almost all expressivist solutions to the Frege-Geach rely on a property of attitudinal discordance, the question arises of whether this is a property of the *content* of the attitudes or, rather, a property of the attitudes themselves.

- **A-type discordance** A set of attitudes is A-type discordant if and only if it is discordant in virtue of being a set of attitudes of the same type with inconsistent contents.

- **B-type discordance** A set of attitudes is B-type discordant if and only if it is discordant but not in virtue of being a set of attitudes of the same type with inconsistent contents.

The contrast between these cases is clearest in the distinction between cases like belief and cases like approval and disapproval—a pair of beliefs is discordant in virtue of having logically inconsistent contents. But the discordance between disapproval and approval does not obviously have the same character—I can approve of two inconsistent contents or actions without discordance, whereas ap-

proving and disapproving of one and the same content or action is clearly discordant (Baker and Woods p. 409). The discordance between approving and disapproving of the same action is, at least *prima facie*, due to features of the attitudes themselves, not the action which is their content. The A-type theorist holds that all attitudinal discordance—or, anyways, all attitudinal discordance relevant for the analysis of normative thought and talk—strongly resembles belief. So if disapproving of murder and tolerating it are discordant, that's because of a subtle feature of the content. The B-type theorist denies this, holding that the appearance of B-type discordance matches the reality.

Motivating a notion of discordance, whether A- or B-type, does not by itself solve the Frege-Geach problem. We use discordance to characterize expressed mental states in such a way as to (a) license the right implicational properties between statements and (b) distinguish expressed mental states of sentences obviously different in meaning. If we can specify the action of the logical connectives in such a way that complex statements which would ordinarily be inconsistent turn out to express discordant mental states, then we have come a long way towards giving a *constructive* semantics in the sense in which the minimalist and inferentialist solutions seem to fail to. Even if one disagrees that giving a constructive semantics is necessary to solve the Frege-Geach, it is clearly the *best* type of solution.

Discordance solutions, prior to Schroeder's *Being For*, have fallen into a few determinate B-type accounts: higher-order solutions and commitment accounts (due mainly to Simon Blackburn), and norm-acceptance or planning-state solutions (due to Allen Gibbard). We will now turn to describing how these work in broad strokes before turning to A-type accounts—Schroeder's *being for* proposal and hybrid expressivism. We will close by drawing a few lessons for forward progress on the problem.

B-type Discordance Theories: Higher-order Attitudes and Commitments

Higher-order attitude solutions were first developed in Blackburn's *Spreading the Word* (1985). They treat logical operators—and presumably cognate expressions—as indicating expression of attitudes towards holding viewpoints consisting of patterns of attitudes. Asserting "murder is wrong" expresses *disapproval* of murder (DIS[murder]) whereas asserting a conditional like "if murder is wrong, then stealing is wrong" expresses *disapproval* of simultaneously disapproving of mur-

der while not disapproving of stealing:

DIS[disapproving of murder while not disapproving of stealing]

In the simple case we *express* an attitude towards murder (as indicated by capital letters), whereas in the embedded case, we *describe* this psychological state. The descriptive content of the embedded sentence is still clearly derivative from the embedded use—it *describes* possession of the attitude a sincere assertion of "murder is wrong" would express.

Suppose we disapprove of murdering and hold that stealing is wrong if murder is. We recapture the mistake in failing to conclude stealing is wrong—failing to disapprove of stealing—by noting that we would then:

DIS[disapproving of murder while not disapproving of stealing]

DIS[murder], not DIS[stealing]

But this means we would self-disapprove, resulting in a sort of fracture in our psychologies. Other cases of valid argument patterns can be treated analogously. Of course, there are problems involving refraining from approving or disapproving and the psychological fracture account requires implausibly treating mixed cases as expressing disapproval of combinations of cognitive and conative states. But put these relatively minor problems to the side for now.

Blackburn's solution is a type of B-type solution—viewpoints containing certain patterns of attitudes, such as tolerating and disapproving of murder, are 'fractured', which is Blackburn's notion of discordance. But discordance, so interpreted, isn't sufficient to explain our reasons to avoid such states. Blackburn claims these reasons are derivative from our *disapproval* of fractured stances. However, this suggests a deep problem for views which explain our reasons to avoid discordance in terms of disapproving of it: we might not always so-disapprove. In fact, given work by Gil Harman and others on the benign character of much contradiction, we may not disapprove of some contradictory belief-states at all. Believing my viewpoint is inconsistent does not trigger self-loathing in me; I view it as the cost of doing epistemic business while maintaining a healthy skepticism about whether I have done it correctly. Moreover, nothing cleanly separates disapproving of stances like the above from disapproving of viewpoints involving approval of, say, anything involving chickens. But, presumably, the types of discordance involved in the two cases are significantly different.

11

Blackburn's solution also seems to conflate *logical* and *rational* or *practical* inconsistency (Van Roojen 1996) "It's wrong to disapprove of murder and not disapprove of stealing" expresses *disapproval* of simultaneously disapproving of murder and not disapproving of stealing, just like the conditional "If murder is wrong, so is stealing." On Blackburn's account, the validity of the above argument arises from the "fractured" nature of the mental states involved in holding the premises while refraining from holding the conclusion. But this means that

P1 Murder is wrong
P2 It's wrong to disapprove of murder and not disapprove of stealing
 C Stealing is wrong

is also valid. Which it is clearly not (though see (Weintraub 2011) and (Baker and Woods 2015) for ways to address this problem.)

Blackburn, in response, started treating normative assertions as expressing *commitment states*—a type of psychological state—to accept certain patterns of attitudes (Blackburn 1988). The idea is that accepting a disjunction ordinarily *commits* you to the truth of at least one disjunct, but it does not commit you to taking either of the particular disjuncts as true. Blackburn so treats an asserted normative disjunction—murder is wrong or stealing is wrong—as expressing a *commitment*:

disapprove of murder or disapprove of stealing

But, importantly, not as expressing commitment to disapproving of either individually. In combination with, say, approval of stealing and thus inability to meet my commitment by satisfying the right disjunct, I am derivatively committed to disapproving of murder on pain of a fractured mental state. Of course, there are many ways to not disapprove of stealing—avoiding taking a stand on it, for instance— and in such cases we have a *rational* commitment to come to disapprove of murdering absent changing our mind.

Since a trivial normal-form theorem guarantees that every sentence of propositional logic can be expressed in terms of negation and disjunction, Blackburn's account can be used to characterize a wide swath of normative and mixed claims. But this is a slightly strange strategy; giving a semantics for a language in terms of a set of logical equivalents transforms the expressivist project from one of *describing* our use of normative language to one of *replacing* or *rationally reinterpreting*

normative language. The usual goal of expressivism is to *interpret* ordinary usage and, for that, providing truth-conditions merely for disjunction and negation is plainly inadequate; our language is richer than that.[2]

Blackburn's move neatly avoids Van Roojen's objection since we no longer interate the notion of disapproval. 'It's wrong to drink or drive' expresses a commitment to disapproving of drinking or driving while 'it's wrong to drink or it's wrong to drive' expresses the commitment to either disapproving of drinking or disapproving of driving, and 'it's wrong to approve of drinking and approve of driving' expresses a commitment to disapproving of approving of drinking while approving of driving. Notice also that it easily solves the case of mixed constructions—we need only treat a disjunction like "murder is wrong or stealing is fun" as expressing a commitment to come to disapprove of murder if we fail to believe that stealing is fun and vice versa. Problems solved.

Or are they? Mark Schroeder and others have argued that such solutions, even though they *formally* solve the Frege-Geach problem, do so by assuming something that needs explaining. Consider the analogous theory of Horgan and Timmons (2006). They posit a pair of commitment operators (is- and ought-commitment) and claim that the negation of a commitment state φ, $\neg\varphi$, indicates a distinct commitment state logically inconsistent with ϕ. This does little to explain what these states are and *why* they are discordant with each other. Since the stipulated 'logical inconsistency' that arises merely from negation isn't, by itself, obviously discordant, they assume a notion of discordance arising from certain combinations of states, mental and otherwise, which needs to be somehow reduced, explained, or made analogous to more familiar notions discordance of belief. According to Schroeder, Blackburn likewise needs to provide an explanation of why committing to approving and disapproving of the same thing is discordant. We will return to this shortly in the context of the most popular solution to the Frege-Geach, Gibbard's factual-normative semantics.

---

[2]Note that the *negation* problem, below, could be similarly "finessed". Another equi-trivial normal form theorem guarantees that every sentence of propositional logic is logically equivalent to a sentence either of the form ϕ or $\neg\phi$ where ϕ contains no negations. This does nothing to undermine the intuitive force of the negation problem; natural language constructions contain a rich array of negative particles.

B-type discordance theories: Plan-laden semantics

Gibbard's factual-normative semantics, Gibbard (1992, 2003), uses a modification of possible-worlds semantics to account for the *semantic values* of moral language. Let a *fact-prac* world be a pair of a set of facts (the worldly component) and a plan about what to do in situations like those described by the worldly component (the practical component).[3] We can represent a claim, possibly including both practical and normative material, as a set of fact-prac worlds—the set of fact-prac worlds which are consistent with the plan described by the normative claim and the content of the descriptive claim.

Consider, for example, the claim that murder is wrong and stealing is fun. Its content is the intersection (green) of the set of fact-prac worlds where we plan to blame for murdering (yellow) and the set of fact-prac worlds where stealing is fun (blue). That is, its content is $\{@\}$.

	S is F	S is not F
P-to-B for M	@	w
P-to-not-B for M	u	v

Accepting this claim means intersecting this set with the set of fact-prac worlds we haven't yet ruled out, typically resulting in a smaller set of worlds. This "semantics" models both normative and descriptive content analogously to the way ordinary (coarse-grained) linguistic semantics models descriptive content.

Of course, a formal semantics is not yet, by itself, an explanation of meaning (Burgess 2008). In order to bridge the gap between an admittedly appealing formal model of meanings and meaning itself, we need to interpret the formal semantics. Gibbard does this by taking the state-of-mind expressed by a set of fact-prac worlds to be a *planning state*—a state of planning to do such and so depending on the circumstances that obtain. Or, in a later epicycle, by additionally taking the *complement* of a set of worlds Γ—the set of all fact-prac worlds which are not in Γ—as modeling the mental state which disagrees with the mental state indicated by Γ.

---

[3]Gibbard initially characterized the practical component in terms of the acceptance of norms for feeling and behavior. In his later work, he used planning states to explicate his earlier picture. The differences between these approaches aren't especially important here, so we set them aside.

Our example of disjunction, for examples, expresses that our settled-with-respect-to-every-possible-circumstance—or, in Gibbard's lingo, *hyperdecided*—state of mind is one where we plan to blame for murdering if it ain't fun. Accepting the disjunction cuts away those hyperplans where murder isn't fun, yet we don't plan to blame for murdering—which would be the psychological state which disagrees with planning to blame for murdering if murdering isn't fun. Complexities emerge in using this framework to account for the distinction between indifference—blame for murdering or not, whatever!—and indecision—to blame for murder or not to blame for murder, that is the question! (Dreier 2006), but such complexities can be finessed. Gibbard's interpretation is rightly regarded as one of the most flexible and interesting attempts to solve the Frege-Geach problem.

It is so flexible that there have been attempts accept something like Gibbard's *formal* framework, but use it to model a distinct interpretation of normative thought and talk. Nate Charlow (2013), for example, uses roughly the Gibbardian framework to give a *semantics* for normative language which doesn't involve mental states at all, even though he accepts the *metasemantic claim* that expressed mental states ground the meaning of normative claims. (Carballo 2014) claims that expressivists can all accept something like Gibbard's framework as a mediate *semantic* story between a standard linguistic semantics and the expressivistic metasemantic interpretation thereof. The Gibbardian framework, if it can be properly interpreted, seems to be a very promising approach to solving the Frege-Geach.

Which is not to say there aren't problems. Schroeder (2008) also charges Gibbard with taking for granted that which needs to be explained—the notion of discordance or mental fragmentation—which does the important work of rescuing the intuitive notion of inconsistency for the expressivist. To see more clearly what the objection is, note that we've interpreted the mental state expressed by 'Murder is wrong' as the set of hyperplans where we plan **to blame** *for murdering*. This two-part structure—X-ing for Y-ing—solves a problem involving negation we will see in a minute. But Gibbard's basic normative notions are not "wrong" or "permitted", but "the thing to do" or "ought" in a generic sense, whose interpretation is simpler. 'We ought to murder', in this sense, simply expresses planning to murder. And this gives rise to a problem—formulated originally in (Unwin 1999)—which has been recently highlighted by Schroeder in one of the most trenchant discussions of the Frege-Geach problem to date.

The Negation Problem

Consider three ways of inserting a negation into normative judgment:

1a Jack thinks murdering is not the thing to do.

2a Jack thinks that not murdering is the thing to do.

3a It's not the case that Jack thinks murdering is the thing to do.

Planning states, at least in their toy implementation, only capture two of these three. Not planning to murder analyzes 3a, planning on not murdering gets us 2a, but what gets us 1a? Similar troubles arise for a Blackburn-style solution—not disapproving of murdering corresponds to not thinking that murder is wrong, disapproving of not murdering analyzes thinking that not murdering is wrong, but thinking that murdering is not wrong sits unanalyzed. Any solution to the Frege-Geach problem which fails to account for these three natural language interpretations will be extensionally inadequate. Failure to solve the negation problem thus implies failure to solve the Frege-Geach problem.

Blackburn solves this problem by treating a construction like 'thinking that murder is not wrong' as expressing something like *tolerance*—the minimal state conflicting with disapproval—of not murdering. Gibbard, by his assumption that we are dealing with *complete plans*, collapses 1a and 3a since if my complete plan doesn't include murdering, then it must include not murdering—though one might worry about Gibbard's ability to deal with more psychologically realistic *incomplete* plans.

Schroeder grants that these types of solutions would work, but worries that they take for granted unexplained notions of discordance. Against Blackburn, he claims it is unclear why tolerance and disapproval of one and the same thing are discordant. Against Gibbard, Schroeder claims that it is unclear that there exists a mental-state corresponding to sets of hyperplans which disagree with a state like planning to murder (note, though, that this is a problem with assigning mental states to formal objects, not a direct problem with the notion of discordance itself.)

Many have followed Schroeder in this (Charlow 2013, etc), even though attitudes which are fundamentally incoherent with one another is a familiar phenomenon once we move outside of the domain of *propositional* attitudes (Baker

and Woods 2015). Explaining B-type discordance likewise does not seem impossible as "...it is arguable that the root idea of inconsistency is precisely the idea of disagreeing with oneself" (Wedgwood 2009). Putting these B-oriented responses to the side for now, Schroeder is right that expressivists need to justify that a *constructive* approach to expressivist semantics is possible. In his *Being For* (2008), he shows what an A-type constructive approach would look like.

A-type discordance theories: Being For

Schroeder starts by stipulating a single attitude—*being for*—that is inconsistency transmitting. To avoid confusion, we will use *discordant* for the relevant property of attitudes and *inconsistent* for the semantic property of their contents. Then:

> An attitude A is *inconsistency-transmitting* just in case two instances of A are discordant when their contents are (semantically) inconsistent.

The idea is to see how far an expressivist theory can get if we just *assume* that there is an attitude with the right sorts of discordance-properties. It is thus very important to keep in mind that *being for* is a placeholder for some existing attitude to be filled in later (if the development gets that far!) Since Schroeder has argued against B-type theories and rejects minimalist ones, he views his account as the only way for pure, not hybrid, expressivism to proceed.

Given that *being for* is, by stipulation, an inconsistency-transmitting attitude, Schroeder implements Gibbard's solution of treating a judgment of, say, wrongness in terms of the structure [X-ing for Y-ing], yielding something like [being for blaming for murdering]. We can then interpret

4a Murder is not wrong.

5a Not murdering is wrong.

6a It's not the case that murdering is wrong.

as

4b being for not blaming for murdering.

5b being for blaming for not murdering.

6b not being for blaming for murdering.

Since, unlike planning attitudes, there is no reason to assume that there is a linguistic expression for attitudes like *being for drinking*, we can neatly skirt the worry about solving the negation problem for sentences expressing one-part attitudes (though see (Gibbard 2014, Appendix 2) for Gibbard's response.)

Schroeder extends his basic interpretation to a compositional semantics for the logical operators, yielding a flexible and interesting variation on the usual sort of solution to the Frege-Geach. His view has the virtue of modeling what is expressed as almost directly analogous to the case of belief where, as Schroeder notes, we have very good reason to think that inconsistency transmission holds— even though there yet is no compelling *explanation* of *why* it does. Unfortunately, the initial solution has problems accounting for mixed expressions since if what is expressed by 'murder is wrong' is being for blaming for murdering and what is expressed by 'murder is fun' is the belief that murder is fun, then we need to find some state which is expressed by 'murder is wrong or murder is fun'. It certainly isn't a belief on pain of giving up non-cognitivism.

Schroeder, in response, goes expressivist about belief as well—asserting 'grass is green' expresses being for proceeding as if grass is green—but the resulting view itself has significant problems. Schroeder puts forward a solution using what he calls *biforcated* attitude semantics—pairs of being-for attitudes such as

being for proceeding as if grass is green

being for not proceeding as if grass is not green

but the overall view is clunky, requiring special pleading for each additional linguistic construction in order to accommodate iterated versions of the negation problem (violating one of our desiderata for a solution to the Frege-Geach.) As Schroeder also notes, his view also has trouble with modal constructions and other complicated parts of natural language. The details are complicated and would distract here (but see ch. 8-12 of *Being For*.) It should be noted that Schroeder is one of the only theorists to have even attempted to develop an expressivist semantics to this extent; most theorists stop after accounting for a logical operator or two.

On balance, if Schroeder's objections to B-type expressivism succeed and if his A-type view really is the only way to go, then expressivism and other views

subject to the Frege-Geach objection are in significant trouble. But, luckily, this is plausibly not the case: none of the problems for B-type expressivism is conclusive and B-type accounts are at least superficially more descriptively adequate as an account of what normative language expresses (Baker and Woods 2015, Wedgwood 2009). Schroeder's view, on the other hand, is far more developed than many alternatives and is a flexible and useful alternative to the more usual B-type accounts. The outcome of this internecine dispute awaits even more serious development of both. We now turn, finally, to the other extant A-type solution to the Frege-Geach problem—hybrid expressivism.

A-type Discordance Theories: Hybrid Expressivism

Hybrid expressivists claim that assertions of normative sentences express *both* descriptive beliefs and conative states. These views easily solve the Frege-Geach problem as long as the belief component is importantly related to the conative component. For example, consider a hybrid view where an assertion of "murder is wrong" expresses the belief that murder is G—where G is some descriptive property of actions—and disapproval of things which are G. The belief component accounts for embedding and fixes the meaning-conditions of complex embeddings. For example, we can treat "if murder is wrong, stealing is wrong" as expressing the belief that if murder is G, then stealing is G *and* disapproval of G things. Likewise for more complex embeddings.

One might worry here that there is no property G which all linguistically competent speakers ascribe to wrong things (Schroeder 2009). But a more general solution can be developed where an assertion like "murder is wrong" expresses a single hybrid or relational higher-order state composed of a descriptive component like murder is G and disapproval of G things—allowing that the G might be filled in by different properties (Schroeder 2013, Ridge 2014, Toppinen 2014). Perhaps when I say "murder is wrong", I express a state consisting, descriptively, of the claim that murder is in violation of the rights of others and disapproval of things which violate the rights of others, whereas when Richard says "murder is wrong", he expresses a functionally identical state consisting of the claim that murder is non-utility-maximizing and disapproval of things which don't maximize utility. The solution to the Frege-Geach problem on these more sophisticated accounts is analogous to the above; logical operators and other semantic operators are applied to the cognitive aspect of the higher-order state, leaving the conative portion "scoped out" over the entire complex judgment.

Since inconsistency properties are all foisted off on the expressed belief (or the belief component of a complex state), hybrid expressivism neatly solves the Frege-Geach problem. There are other worries for this sort of account; for one thing, so much is foisted off on belief that it becomes unclear in what sense it preserves the initial appeal of non-cognitivist views. It is also not obvious that the resulting views are preferable to, say, cognitivist views asserting a normative claim *expresses* a belief and *implicates* some conative state. Views which accept the expressivist's claim about the connection between normative assertion and conative states, but claim the connection is part of the conversational *pragmatics* of normative discourse have the advantage of fitting more closely with contemporary semantic views, after all (See Finlay 2014 for sophisticated development of this point).

The contemporary landscape

Where do things stand after our brief survey of solutions to the Frege-Geach problem? As the above demonstrates, there are a number of potential ways to solve the Frege-Geach problem. Schroeder's argument against B-type expressivism isn't conclusive—he establishes that B-type theorists face the additional burden of explaining discordance attaching to their preferred mental states *as well as* explaining discordance of belief, but not that doing so is unworkable. Dreier's problem of distinguishing between indifference and indecision seems closer to a research problem than a devastating objection. And even van Roojen's problem points towards the burden of distinguishing logical discordance from pragmatic and semantic discordance, not the inability to do so (see (Baker and Woods 2015, §4.))

This is fertile ground for abductive argument. Consider weighing Gibbard's view against Schroeder's *being for* approach. Each has fairly straightforward costs—Schroeder's approach requires us to go non-cognitivist about descriptive discourse, Gibbard's approach is developed for idealized agents and thereby has trouble interpreting finite and mildly irrational agents like ourselves—but each has significant advantages as well. Whether or not to adopt one or other of these approaches is a matter of weighing out the various costs and benefits of so doing so.

But at this point it is crucial to enjoin caution. Expressivist semantics, on whichever approach, have barely left their infancy. There is simply no developed

comparison of expressivist accounts to the truth-or whatever-conditional accounts of the sort taken for granted in contemporary linguistic semantics. Given this situation, it is crucial to not conflate the property of not *yet* being able to solve a problem—such as interpreting modal constructions or accommodation of a sophisticated theory of subjunctive conditionals—with the property of being *unable* to solve a problem. What developments sustained development of an expressivist semantics would bring is, as yet, mostly a mystery. Even a quick glance at the theoretical maneuvers sketched above demonstrates that vastly more can be done with expressivist semantics than one might have thought at the outset.

Related Topics

Expressivism, Thick Concepts, Hybrid Expressivism, Inferential Role Semantics and Metaethics, Cognitivism and Non-Cognitivism, Metasemantics and Metaethics, Revisionism vs. Conservativism in Metaethics

Acknowledgments

References

Alwood, A. 2015. Should Expressivism Be a Theory at the Level of Metasemantics? *Thought: A Journal of Philosophy*.

Baker, D. and J. Woods. 2015. How Expressivists Can and Should Explain Inconsistency. *Ethics 125*(2): 391–424.

Blackburn, S. 1985. *Spreading the Word*. Oxford University Press.

Blackburn, S. 1988. Attitudes and Contents. *Ethics*: 501–517.

Burgess, J. 2008. Tarski's Tort.

Carballo, A. P. 2014. Semantic Hermeneutics. *Metasemantics: New Essays on the Foundations of Meaning*: 119.

Charlow, N. 2014. The Problem with the Frege-Geach Problem. *Philosophical Studies 167*(3): 635–665.

Chrisman, M. 2015. *The Meaning of 'Ought': Beyond Descriptivism and Expressivism in Metaethics*. Oxford University Press.

Dreier, J. 1996. Expressivist Embeddings and Minimalist Truth. *Philosophical Studies 83*(1): 29–51.

Dreier, J. 2006. Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution. *Oxford Studies in Metaethics* 1: 217–233.

Finlay, S. *Confusion of Tongues: A Theory of Normative Language*. Oxford University Press.

Frege, G. 1956. The Thought: A Logical Inquiry. *Mind*: 289–311.

Geach, P. T. 1960. Ascriptivism. *The Philosophical Review*: 221–225.

Gibbard, A. 1992. *Wise Choices, Apt Feelings*. Harvard University Press.

Gibbard, A. 2003. *Thinking How to Live*. Harvard University Press.

Gibbard, A. 2014. *Meaning and Normativity*. Oxford University Press.

Grice, H. P. 1957. Meaning. *The Philosophical Review*: 377–388.

Hare, R. M. 1952. *The Language of Morals*. Clarendon Press, Oxford.

Hare, R. M. 1971. *Practical Inferences*. Cambridge University Press.

Horgan, T. and M. Timmons. 2006. Cognitivist Expressivism. In *Metaethics After Moore*, 255–298. Oxford University Press.

Ridge, M. 2014. *Impassioned Belief*. Oxford University Press.

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

Schroeder, M. 2009. Hybrid Expressivism: Virtues and Vices. *Ethics 119*(2): 257–309.

Schroeder, M. 2013. Tempered Expressivism. *Oxford Studies in Metaethics* 8: 283–314.

Searle, J. R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Sinclair, N. 2011. Moral Expressivism and Sentential Negation. *Philosophical studies 152*(3): 385–411.

Skorupski, J. 2012. The Frege–Geach objection to expressivism: still unanswered. *Analysis 72*(1): 9–18.

Stoljar, D. 1993. Emotivism and Truth Conditions. *Philosophical Studies 70*(1): 81–101.

Toppinen, T. 2013. Believing in Expressivism. *Oxford Studies in Metaethics* 8: 252–282.

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

Van Roojen, M. 1996. Expressivism and Irrationality. *The Philosophical Review*: 311–335.

Wedgwood, R. 2010. Schroeder on Expressivism: For–or Against? *Analysis 70*(1): 117–129.

Weintraub, R. 2011. Logic for Expressivists. *Australasian Journal of Philosophy 89*(4): 601–616.