Moral inferentialism and the Frege-Geach problem

Mark Douglas Warren

Abstract Despite its many advantages as a metaethical theory, moral expressivism faces difficulties as a semantic theory of the meaning of moral claims, an issue underscored by the notorious Frege-Geach problem. I consider a distinct metaethical view, inferentialism, which like expressivism rejects a representational account of meaning, but unlike expressivism explains meaning in terms of inferential role instead of expressive function. Drawing on Michael Williams' recent work on inferential theories of meaning, I argue that an appropriate understanding of the pragmatic role of moral discourse—the facilitation of coordinated social behavior—suggests the kind of inferences we should expect terms with this function to license. I offer a sketch of the inferential roles the moral 'ought' plays, and argue that if we accept that the relevant inferential roles are meaning-constitutive, we will be in a position to solve the Frege-Geach problem. Such an inferentialist solution has advantages over those forwarded by expressivists such as Blackburn and Gibbard. First, it offers a more straightforward explanation of the meaning of moral terms. It also gives simple answers to at least two semantic worries that have vexed contemporary expressivists—the "problem of permissions" and the commitment to "mentalism", both of which I argue are problems that don't get traction with an inferentialist approach. I conclude by considering ways in which this approach can be expanded into a more robust semantic account.

M. D. Warren (✉)
Department of Philosophy, Miami University, 212 Hall Auditorium, Oxford, OH 45056, USA
e-mail: MarkDouglasWarren@gmail.com

1 Introduction

Expressivism about moral discourse comes with a host of well-known advantages. Perhaps most importantly, it offers an account of moral discourse that sits comfortably with naturalism—it is an account that does not rely on any queer non-natural moral facts that carry with them a mysterious intrinsic property of to-be-doneness (Mackie 1977). Nor does it rely on a dubious reduction or identification of moral facts with natural facts. Expressivists argue that because moral language functions primarily to express mental states—and not to report on a domain of moral facts—we can give a naturalistically respectable explanation of moral discourse that doesn't appeal to any problematic moral truth-makers.

But the basic semantic approach that expressivists use makes them vulnerable to the notorious Frege-Geach problem: how can expressivists account for the uniform content of moral terms even when they are uttered in "force-stripping" contexts, where their typical use is not in effect? Though there are several attempted responses to this issue (Blackburn 1984, 1988, 2006; Gibbard 1990; Horgan and Timmons 2006), recent literature on the topic suggests that these are deeply problematic (Unwin 1999, 2001; Schroeder 2008a; van Roojen 1996; Schueler 1988; Merli 2008), and leave the expressivist with a heavy explanatory burden, most of which involves explaining exactly how mental states are supposed to inform the meaning of moral claims. In this paper, I will argue that there is an alternative position, moral inferentialism, can sidestep this burden but that nevertheless retains expressivism's core advantages. For this reason, it's a position worthy of our attention.

After giving a brief sketch of metaethical inferentialism in Sect. 2, I explain the Frege-Geach problem and how expressivists deal with it in Sect. 3. In Sect. 4, I summarize Williams' meta-theoretical framework for explaining the meaning of a term by way of articulating the inferential rules that enable its functional role in a particular discourse. I argue in Sect. 5 that the characteristic function of moral discourse is its role in encouraging coordinated social behavior. This is the function that makes sense of the inferential rules that govern our moral terms; in Sect. 6, I offer a sketch of the meaning constitutive inferential rules that govern the moral term 'ought', and argue that these are the ones we should expect a moral term to have, if they are to enable coordinated behavior. In Sect. 7, I'll argue that such inferentialism enjoys distinct advantages over expressivism: it is a simpler position that can solve the Frege-Geach problem without recourse to the mental states on which its expressivist cousin relies. This means that inferentialism can avoid many of the semantic complications that come in the wake of an expressivist solution to the Frege-Geach problem. In Sects. 8 and 9, I'll consider two well-known complications from the literature, and show how an inferentialist account avoids them. I conclude that inferentialism is a metaethical position that preserves the advantages of expressivism while avoiding some of its more serious problems, and that it thus deserves further development and investigation.

2 Inferentialism: expressivism without expressing

Metaethical inferentialism is a very close cousin to expressivism; both can rightly be understood as grandchildren of Ayer's (1936) emotivism. On Ayer's conception, moral claims lack "factual meaning" (Ibid, p. 107); moral claims do not correspond to anything that can be considered true or false. Instead, they simply evince one's approval or disapproval. For this reason, Ayer's account has been disparagingly called the "Boo/Hoorah" theory of ethical claims, since he holds that ethical claims such as, "Murder is wrong" have no content more than the interjection, "Boo murder!" This approach is characteristic of both inferentialism and expressivism in two ways: first, it denies that moral assertions function to represent moral facts, properties, or states of affairs. In this sense, emotivism, expressivism, and inferentialism are what Huw Price would call *non-representational* accounts of metaethics. This is in contrast to representational approaches, which assume that once we've characterized the role of terms on the 'near end' of the word-world semantic relationship inherent in moral claims (i.e., our use of normative terms such as 'good' and 'ought'), we must then go on to characterize the special moral truthmakers for such claims—those properties or relations that lie on the 'far end' of this semantic relationship.[1] The non-representational approach rejects this assumption; if we can explain the features of moral discourse *pragmatically*, without involving the referents of terms in any explanatory role, we won't need to engage in this sort of truthmaker-hunting project; all the explanatory work will have been done.

Second, emotivism doesn't take this to entail an error theory about morality; instead, the approach is to look at moral discourse itself in a new way. Again, according to Ayer, moral statements are essentially evaluative, and not truth-apt. Because the question of a moral statement's truth shouldn't come up, we have no reason to worry that moral statements do not correspond to any moral facts in the world.

But such an emotivist approach does boil down to a radical reformation of ethical discourse. We think we can speak non-problematically about moral beliefs and moral errors, but because moral claims have no content, the structure of ethical

---

[1] See Price and Macarthur (2007, especially pp. 94–97) for more on the distinction between representational and non-representational approaches. For examples of representational approaches to metaethics, see Boyd (1988); Brink (1989); Brandt (1979); Smith (1994); Sturgeon (1988). See Timmons (1999, Ch. 1) for the role that the representational assumption has played in modern metaethics:

> [An] operative assumption regarding matters ontological was that they were to be handled linguistically—that questions about whether there are moral properties or facts and, if so, whether they are identical to non-moral properties and facts were to be answered by settling questions about the meanings of moral terms and expressions. (Ibid., p. 26)

Timmons' argues that metaethicists who reached radically different ontological conclusions also undertook this same fundamentally representational approach. For example, Mackie (1977) accepts "a certain [representational] account of the meaning of moral terms and expressions, according to which ordinary moral statements purport to be about objective moral properties or facts, but then [denies] that there are any such properties or facts." (Timmons 2006, p. 27) The same interpretation is available for relativistic accounts like those found in Wong (1984) and Prinz (2007).

discourse is grammatically and logically misleading, and we'd do well to look past its surface features. This is a high price to pay to be an emotivist: if Ayer is right, it's hard to make sense of some very basic features of moral talk in a way that doesn't deeply undermine the discourse. The denial that ethical claims have meaning seems to undermine important ethical notions like moral error, disagreement, and improvement. Expressivism avoids this conclusion by arguing that ethical claims do in fact have content, while denying that such content is given in terms of *representation*. It retains Ayer's basic contention—that moral claims function to express rather than describe—but argues that it is this function that actually gives moral claims their meaning. Expressivists explain the semantic features of moral claims by reference to this essentially expressive function; the meaning of moral claims is explained in terms of the mental states they are used to express.

If we understand the progression from emotivism to contemporary expressivism in this way, metaethical inferentialism can be characterized with a further distinction. It retains the non-representational aspect of expressivism's semantic account, but denies that the fundamental non-representational function of moral language is best understood in terms of the mental states moral claims serve to express. Eschewing the expression relation as semantically fundamental, the inferentialist understands "the commitments which constitute a claim's meaning not in ontological or psychological terms but in inferential terms". (Chrisman 2010, p. 118) The inferentialist argues that we can identify the meaning of moral claims by articulating their inferential connections to other claims, and that these connections are best understood by looking to the pragmatic function moral discourse fulfills.[2]

This development should not be such a surprising step in the evolution away from emotivism, if for a moment we look away from debates in metaethics and cast our

---

[2] As a semantic theory, inferentialism faces well-known objections. See for example Williamson (2003, 2007). Williamson acknowledges that these objections apply only to *naturalist* versions of inferentialism (e.g., Block 1986; Field 1977; Harman 1999; Horwich 1998; Loar 1981; Peacocke 1992), according to which it is either speakers' actual use of concepts, or their dispositions to employ concepts in certain ways that establish the meaning-constitutive rules that govern our concepts. Williamson admits (2003, p. 291), though, that these objections don't apply to *normativist* versions of inferentialism (such as those developed in Brandom 1994, Gibbard 1994), according to which the meaning or our concepts is given by the inferences we *ought* to make with them. See Thomasson (2014) for a normativist reply to Williamson. Obviously the prospects for metaethical inferentialism will depend on the viability of some form of inferentialism as a general semantic theory, but a defense of this semantic theory, and investigation into its best form, lie beyond the scope of this paper. To those who are skeptical about inferentialism, the results here may be seen as conditional: if inferentialism is a defensible semantic theory, then it makes available a plausible and interesting account of moral discourse.

In any case, the expressivist shouldn't take too much comfort from the attacks on inferentialism: like inferentialists, expressivists attempt a non-representational account of moral thought and discourse by considering a linguistic *explanans* (in the case of the expressivist: mental states. In the case of the inferentialist: the terms themselves) and the inferential profile these have. It seems plausible that any objection that fells inferentialism as a general semantic theory will spell trouble for the expressivist also. So for example, it's possible to read some contemporary criticisms of expressivism (e.g., van Roojen 1996; Merli 2008) as extensions of Williamson's basic attack on inferentialism: Williamson's principle complaint against inferentialism is that one can competently employ a concept without standing ready to engage in the sorts of inferences that concept would license. This is quite similar to the criticisms delivered against expressivism: it seems one can make certain moral judgments without being in the mental state those judgments are supposed to express.

eye on non-representational accounts of other types of discourse. For example, Paul Horwich argues that use of the truth predicate in assertions is important only as a generalizing device; the meaning of "P is true" is given by the equivalence schema: the proposition expressed by 'P' is true iff P. (Horwich 2010). Gilbert Ryle explains the import of law-like scientific claims in terms of the inference-tickets such claims serve to license; a claim such as "Salt dissolves in water" licenses the inference from "X is salt" to "X will dissolve in water" (Ryle 1949). Amie Thomasson argues that modal claims like "Necessarily Q" make explicit the semantic rules that govern our use of the term 'Q'. (Thomasson 2007) Notice that all of these non-representational accounts, like expressivism, involve explanations of the discourses at hand that do not rely on some peculiar moral/alethic/causal/modal range of facts waiting to be described. Each instead explains the discourse in terms of how we *use* the concepts in question. But it is only expressivism that ties this use down to mental states. Viewed this way, this seems a peculiar fact. I suspect that this feature of expressivism can be understood as a result of the contingent fact of its emotivist lineage; inferentialism offers us a new way to understand the uses of moral terms that inform their meaning.

As we will see, one reason to prefer the inferentialist revision of expressivism is to avoid the briar patch of a semantics based on the inferential relationships between mental states. Another reason for the undertaking, though, is that we should have a *prima facie* preference for such an emendation, purely on grounds of simplicity: an explanation that relies on the functional role of moral terms themselves is more straightforward than an explanation that relies on the functional role of the mental states that moral terms characteristically express. The semantic order of explanation for the expressivist starts with mental states that are expressed by moral terms and then spells out the inferential roles the states themselves have;[3] the content of moral claims comes from the attitudes expressed. But the metaethical inferentialist argues, "there is no need to build the expression relation between ethical claims and …states of mind into the *semantic content* of ethical claims." (Bar-On and Chrisman 2009, p. 133) Instead, and more directly, we can look to the inferential role the moral terms themselves have—without worrying about the mental states expressed—in order to give an account of the semantics of moral discourse. We'll have a simpler explanation because we don't need to assume anything about the expression relationship between moral claims and mental states, nor will there be any explanatory pressure to explain just how it is that mental states enter into inferential relationships with one another.

Before we see how such an account might work, it will help to first review the basic semantic worry associated with expressivism, along with the standard response contemporary expressivists offer. We'll then be in a position to appreciate the advantages inferentialism enjoys over expressivism.

---

[3] See for example Blackburn (1998), Horgan and Timmons (2006), and Gibbard (2003), especially chapter 3.

3 The Frege-Geach problem and the standard response

According to the expressivist tradition, the meaning of claims like "It is wrong to lie" are best understood in terms of the expression of some mental state, for example a disapproval of lying.[4] We use such moral claims to express these evaluative states, and this use gives the claims their meaning. But Peter Geach (1965) argued that those who hold that the use of a term or phrase gives its meaning have things backward: we must first have an understanding of meaning *before* we can give an account of use. This, he argues, is the only way to make sense of conditionals with embedded moral clauses. So, for example, in asserting, "If it is wrong to lie, it is wrong to get your brother to lie", one is neither expressing one's disapproval of lying or of getting your brother to lie; what is asserted is that a conditional relationship holds between these two propositions. The Frege-Geach problem was originally presented by way of conditionals, but of course we find examples of such "force-stripped" clauses not only in conditionals, but also in moral expressions like:

"Is lying wrong?"
"It's not the case that lying is wrong."
"Lying is either harmless or it's wrong."

If the meaning of the word "wrong" in these cases can't be understood in terms of the speaker's disapproval of lying, the expressivist must give some other explanation. If we try to sidestep this argument by claiming that the meaning of moral propositions in such force-stripping conditionals is in fact different from the meaning they have when presented bare, an obvious trap awaits. This sort of response leaves us guilty of *equivocation* in *modus ponens* arguments like the following:

(1) If it's wrong to lie, it's wrong to get your little brother to lie.
(2) It's wrong to lie.
(3) It's wrong to get your little brother to lie.

For this argument to be valid, the meaning of "it's wrong to lie" must be the same in (1) and (3). If it isn't, the argument is invalid because of the equivocation involved.

A few expressivists have tried their hand at responding to this challenge (Blackburn 1984, 1988, 2006; Gibbard 1990; Horgan and Timmons 2006); I want to point to general features these responses have in common. What's needed, it seems, is a compositional semantics that understands the meaning of a complex sentence like (1) given as a function of the meaning of its parts. So the meaning of (1) is a function of the *mental state* expressed by (2) and (3). The implication is that the content of (1) is also given by a mental state being expressed—a complex state, the

---

[4] Expressivists differ on which sort of mental state is expressed by moral claims—approval, intentions, commitments, or plans, etc. For the sake of simplicity I'll simply speak of approval and disapproval.

meaning of which is determined in part by the attitude that *would* be expressed in an force-stripped context by its clauses. (See Schroeder 2008a, pp. 20–21).

"It's wrong to get your little brother to lie" has the same content in (1) as it does in (3), even though (1) does not directly express disapproval of the act, so this sort of response clearly refutes the charge of equivocation. But why does the content of (1) and (2) taken together license the inference to (3)? For an expressivist, the validity of the inference from (1) and (2) to (3) must be explained in terms of the mental states these sentences express. The idea is proposed that it if one has the mental states expressed by both (1) and (2), one is compelled by *force of logic* to also accept the mental state expressed by (3)—to accept (1) and (2) but reject (3) would entail a state of contradiction. The explanation is fundamentally non-cognitive; expressivists don't cash it out in terms of logically inconsistent representations of the world, but instead either specifically in terms of jointly unsatisfiable plans, intentions, evaluations, or other conative mental state,[5] or more generally by pointing out the practical importance of consistency in one's conative attitudes.[6]

This is a tidy sort of response, but it seems to carry with it a host of complications and unintuitive commitments, many of which can be found in the literature on expressivism in the past 20 years (Unwin 1999; van Roojen 1996; Schroeder 2008a; Dreier 2006; Merli 2008; Schueler 1988). For example: It's not clear that the logical notions of consistency and inconsistency are appropriately ascribed to the relationships between non-cognitive mental states (Schueler); an account that explains the semantics of normative claims in terms of evaluative states also seems to involve a conception of validity that entails validity in arguments that common sense would tell us aren't valid (van Roojen); such an account has trouble explaining the logical relations between obligations and permissions (Unwin) and between seemingly contradictory claims (Merli); and perhaps most devastatingly, making normative content dependent on non-cognitive mental states in the way expressivists argue for commits us to also explaining descriptive content in and the meaning of logical connectives in the same way—b which would involve the expressivist in a radical modification of compositional semantics for non-normative language (Schroeder). Obviously, there are arguments and attendant counter-arguments to address these issues (see Dreier 2006; Mabrito 2009; Alwood 2010), but in this paper I want to suggest that one can avoid at least some of these problems by adopting a semantic account that answers the Frege-Geach problem *without* depending on mental states as the fundamental driving force for a non-representational semantic account of moral language.

To accomplish this, I'll focus on the last two from the list of complications above, on the grounds that these have proven to be particularly tricky for the expressivist. I'll explain them in more depth below, but to help keep our eyes on the target, let's name them both:

---

[5] See Horgan and Timmons (2006), Timmons (1999, especially p. 163), Blackburn (1998), and Dreier (2006).

[6] See Blackburn (1984, 2006) and Gibbard (1990).

The Problem of Permissions. Because the expressivist relies on the notion of mental states being in conflict with one another to explain logical contradiction between moral claims, she must account for the apparent contradiction between statements of obligation and permission with reference to a *sui generis* category of mental state—tolerance—that has its own unique semantic properties.

The Commitment to 'Mentalism'. In order to consistently explain logically complex sentences that involve both normative and descriptive atomic parts, the expressivist is committed to *mentalism*, the unusual semantic view that the content of both normative *and* descriptive claims must be explained in terms of the mental states each typically serve to express.

Before I show how inferentialism can meet (or avoid) both of these challenges, though, I'll start by explaining how the view can tackle the basic semantic problem.

4 Meaning from use

To solve the Frege-Geach problem, the task for the moral inferentialist is to give an account of the meaning of moral terms by way of specifying the inferential relationships those terms have, relations that are necessitated by the use to which we put those terms. This account is constrained in the following ways:

C1. It cannot invoke as explanatory a range of moral facts, properties, or relationships that are being represented by moral claims.

C2. This account must explain the meaning of moral terms via the inferential roles that give these terms their meaning, and this meaning—and so too the inferential roles—must remain the same whether the term is embedded or not.

C3. It also has to explain why the meaning of these terms is such that we can use them to make valid inferences—how the conceptual roles played by moral terms informs the inferential relationships we take to be basic in moral reasoning, including the relationships between permissions and obligations, and the validity of argumentative forms like *modus ponens.*

How can we accomplish C2 and C3 without relying on representational resources? We can take a page here from Williams (2010), who develops a meta-theoretical framework to characterize non-representational theories that explain the meaning of terms entirely according to their use. In order to explain the inferential roles that a term plays, he argues, we must first consider the *functional* roles the term plays in our discourse. These will be the roles that such a term must fulfill in order to serve the *pragmatic* role that the discourse plays in human life.[7] It is only once we have such a functional explanation that we can move on to explain the inferential roles

---

[7] Williams himself doesn't distinguish between the pragmatic role of the discourse and the functional role a term in that discourse plays; both are treated in what he calls the "functional component" of an explanation of meaning in terms of use. For the purposes of this paper, it will be helpful to be explicit about the distinction.

the terms associated with the discourse have; these should be the inferential roles—the patterns of usage that competent speakers will be disposed or expected to employ—by which the term fulfills the function of the discourse. These inferential roles determine the content of the term in question. This means that such an account will follow an order of explanation: The *meaning of a term* is constituted by the *inferential roles* that term plays, which are explained by the *functions* that term must fulfill, given the *pragmatic significance* of moral discourse.

Perhaps an illustration is in order. Consider for example a non-representational account from another field: Horwich's (2010) minimalism about truth. According to Horwich, the meaning of the truth predicate can be given by the very simple inferential role given in his *equivalence schema*:

(MT) The proposition *that P* is true if and only if P.

This inferential role is explained by the function of the truth-predicate—it is important as a generalizing device. The truth predicate enables us to simply say, "Everything John said is true", instead of saying, "John said that he's happy, and he is happy; he said that he's coming to the party, and he is coming to the party; and he said that monkeys are funny, and monkeys are funny." We can also use the truth predicate to endorse a proposition even when we're not sure of its actual content: "Chris said something about the transmission needing repair, and I don't know anything about cars. But Chris does, so I'm sure what Chris said is true."[8] Why do we need a term that fulfills such a function? Because of the pragmatic significance of truth-talk: "we need a device that enables us to overcome finite constraints in our effort to describe the world."(Beall 2009, p. 1).

We find in Horwich's minimalism a clear example of how to explain meaning without invoking any sort of underlying property, fact, or relationship. The *meaning* of a term comes from its content-determining *inferential role*: "With respect to giving the meaning of 'true', the rule of use implicit in our acceptance of the instances of MT is explanatorily fundamental." (Williams, p. 322) The inferential role governing our use of the truth predicate is one that enables it to fulfill its *function* as a generalizing device. Having a term that fulfills this function makes sense given the broad *pragmatic import* of truth-talk. Locating the functional role of a term within a practical context "gives the point of our having a word with those [inferential] characteristics." (p. 324).

On the inferentialist view, the semantic rules of use for a term play a constitutive role for that term, similar to the way that the rules of chess play a constitutive role for the chess pieces. So for example what it takes to be a pawn in a game of chess is that it is a piece is governed by certain rules:

- It starts in a certain position—on the second rank of the board.
- It can defeasibly move two spaces the first time you touch it, and only one space otherwise.
- It captures diagonally.

---

[8] For an account of how the inferential rule (MT) explains our use of the truth predicates in contexts like these, see (Horwich 1998, pp. 3–6).

- It promotes to become another piece when you advance it to the eighth rank.
- It can capture *en passant*.

A pawn in chess is in a sense constituted by the rules of the game—if it weren't governed by the rules that say how it can move, it wouldn't count as a pawn. In the same way, the inferentialist argues, the semantic rules governing a term are constitutive of the meaning of that term. To fully understand a game of chess (or the meaning of a term), you have to master these basic rules. But importantly, in chess that doesn't mean that the pawn must display all of these capacities in any given move. As we will see, this gives us an important insight into solving the Frege-Geach problem.

In what follows, I will outline a similar explanation of the meaning of an ethical term—the moral 'ought'. This necessitates a digression on the pragmatic import of moral talk. In the next section, I will look at some of the typical uses of moral discourse—various functional roles it fulfills—and argue that all of these indicate a unified pragmatic role for moral language: the achievement of coordinated social behavior. I will attempt to be as anodyne as possible; the points I make should be recognizable to anyone familiar with expressivist literature.

5 A pragmatic account of moral discourse

Judgments involving moral terms have "indirect and variable influences on action" (Blackburn 1984, p. 189); moral claims typically imply certain motivations on the part of the speaker, and sway the motivations of others. This practical aspect of moral claims is traditionally understood by expressivists in terms of moral judgment internalism, the view that sincere moral judgments necessarily involve some sort of motivation to act in accordance with those judgments. In the interest of skirting controversy, I'll avoid this commitment.[9] For the purpose at hand, there's no need to narrowly construe this feature of moral claims in terms of some necessary connection between judging that 'φ-ing is wrong' and being motivated to not φ-ing. Obviously it's possible to act against one's moral judgments because of fear, lust, greed, and so on—and perhaps one can make moral judgments without any attendant motivational import whatsoever. I'll limit my point to this: there is some sort of conceptual connection between moral claims and influence on actions, however defeasible that connection may be. If we came upon a term in a foreign language that had no such practical import, I suspect we should not translate it as a moral term in our own language. This indicates that an essential feature of moral discourse is its behavioral impact (see Hare 1952, pp. 148–150).

Such impact often comes from the connection between judgments involving moral terms and dispositions to have certain influential emotions: guilt or shame at

---

[9] This hands-off approach also makes sense, given the distinction I'm drawing between inferentialism and expressivism about morality. The expressivist is committed to explaining the meaning of moral claims in terms of the (typically motivating) attitudes they express; inferentialism doesn't carry this commitment, and so can remain silent on whether the characteristic action-guiding feature of moral claims is accomplished in the way that moral internalists argue or by some other means.

having done something one judges wrong, pride at doing right, anger, outrage, contempt, or disgust at those who commit wrongs, and so on (see Gibbard 1990). Notice that these are all emotions that affect social behavior: an aversion to guilt (among other things) keeps one on the straight and narrow; fear of reproof and the punishment or ostracism that is often motivated by moral outrage or disgust can keep even those bereft of moral sentiments from transgressing.

Notice also that our aversion to these emotions isn't sufficient to explain the practical force moral discourse carries. When I tell a child, "You shouldn't lie", I am not offering her a piece of advise on how to avoid punishment or guilt. I do not mean, "You shouldn't lie, unless you're sure you won't get caught and you won't feel bad about it, or unless you simply don't care about being moral." Kant would say that moral claims give us categorical imperatives, since their force (unlike those of hypothetical imperatives) does not depend on an agent having a particular goal or desire. We may wish to avoid all of the theoretical trappings of Kant's account of morality, though, and follow Richard Joyce (2006) in simply ascribing a sort of *practical clout* to moral discourse:

> Whatever kind of practical oomph moral prescriptions are imbued with, it doesn't have its source in internal or external sanctions, nor in some institution's inviolable rules, nor in the desires or goals of the person to whom it is addressed. (Joyce 2006, p. 63)

In short, moral discourse carries with it a kind of social-behavior-influencing force: by making us liable (to others and to ourselves), it influences our behavior with a toolbox of emotional and social incentives and disincentives. The discourse acts as a kind of chaperone for human behavior, constraining selfish or anti-social impulses.

Finally, the practical clout we bring to bear in moral discourse is subject to complicated moral deliberation; we offer our moral reasoning as justification for our behavior and our expectations for the behavior of others (and our expectations regarding the expectations of others), and as challenges to those who disagree. Our moral assertions are offered as objects of discussion for others, and as invitations for potential rebuttal. This capacity to think together on moral issues gives rise to the possibility of reaching consensus, both in moral opinion and the behavior that flows from it. This means that the motivational 'oomph' that characteristically accompanies moral discourse can exert a nuanced influence on our behavior.

Why would creatures like ourselves adopt a discourse with these features? For a highly social, language-using species there is an obvious advantage to employing language that regulates behavior in the ways limned above: it promotes our chances for evolutionary success. Moral discourse offers us a potential counterpoint to the selfish impulses that would drive individuals to uncooperative behavior, even in situations when cooperation would be mutually beneficial; "the evolutionary function of moral judgment [and its attendant discourse] is to provide added motivation in favor of adaptive social behaviors (Joyce 2006, p. 117)." As Mark Timmons argues, we use moral language ultimately to provide.

norms that serve to guide behavior in ways that are crucial to the stability and cohesion of groups of human beings. Because of the importance of certain patterns of behavior (and emotional response), one would expect that humans, capable of complex coordinative social behavior, would evolve more or less shared normative systems that have the sorts of features characteristic of a typical moral outlook. In short, given what human beings are like, the setting in which they find themselves, and that they are capable of quite sophisticated coordinative behavior, one would expect them to develop strategies that would most effectively solve problems of coordination. (Timmons 1999, p. 157)

The purpose of moral discourse, it seems, is to guide us into coordinated social behavior.[10]

This completes the first stage of our explanation of the meaning of moral terms according to their use. The next step, according to Williams' framework, is to provide an inferential account of their meaning that falls out of the above pragmatic considerations. Understanding the inferential role moral terms play is a matter of understanding how terms governed by those inferential roles bring the characteristic behavior-influencing motivational force of moral discourse to bear for the purpose of coordinating social behavior.

6 An inferentialist account of the meaning of 'ought'

In this section, I will sketch an account of the meaning constitutive inferential rules governing our use of the moral 'ought'. The notion of inference here is quite broad (See Brandom 2007). It is not limited to logical inferential connections between

---

[10] This is a rough sketch of features that I believe are essential to understanding what is distinctive about moral discourse, but it is not meant to be exhaustive. There may be other facets that are essential: moral discourse doesn't only goad us into behavior in a particular way; it also seems to goad us towards particular kinds of behavior. There may be certain "*a priori* compulsory propositions that anyone who knows how to use [moral] terms is in a position to recognize as true" (Jackson and Pettit 1995, p. 26). Cuneo and Shafer-Landau (2014) count among these the following:

- It is pro tanto wrong to engage in the recreational slaughter of a fellow person.
- It is pro tanto wrong to break a promise on which another is relying simply for convenience's sake.
- It is pro tanto wrong to humiliate others simply for pleasure. (Ibid., p.7)

I will not consider all of these in depth here, or try to decide which if any are necessary characteristics of moral discourse. For the purposes at hand, it will suffice to echo Allan Gibbard, and point out that these make a particular kind of sense, in light of the broad coordinating function of moral discourse I'm arguing for. Feelings of benevolence prime us to act on behalf of others' interests, and this can lead to coordinated behavior that is to our obvious evolutionary advantage—when it benefits close kin or those in a position to reciprocate, but also because it is to our long term advantage to be a member of a species with altruistic impulses (Gibbard 1990, p. 258). Our concern with fairness lays the grounds for cooperation. "Judgments of fairness stabilize bargaining" (Ibid., p. 262) by giving us a common framework for assessing when gratitude (or retaliation) are appropriate. We might imagine why other issues are routinely within the purview of moral discourse—questions of respect, of disgust and communal purity, of guilt and shame; these are all concerns about our behavior that weigh heavily on our ability to cooperate. And as I argue in Fn. 13 below, if (some or all of) these sorts of considerations are characteristic of the function of moral discourse, we should expect to see these aspects of that function reflected in particular inferential rules.

sentences but also includes material inferences, and supplies us with inferential rules that "take in the circumstances and consequences of application" (p. 654) of terms—what Brandom calls "language exit" and "language entry" (p. 666) moves, respectively—where the circumstances of application may be perceptual rather than propositional, and the consequences may include the practical import of the term's application.

How are we to understand these inferential rules? Recall that for an expressivist, the inferential profile of a moral claim will fall out of the conative mental state that claim is typically used to express. For the inferentialist, though, the inferential rules that govern a term will be cashed out in terms of the social practice of keeping track of how, in asserting a claim with the term in questions, one puts oneself in a position of *liability* and *entitlement*:

> In asserting a claim one not only authorizes further assertions, but commits oneself to vindicate the original claim, showing that one is entitled to make it. Failure to defend one's entitlement to an assertion voids its social significance as inferential warrant for further assertions. (Brandom 1983, p. 641)

So in the case of asserting, "You ought to φ", I make myself liable. I'd be expected to justify my claim in response to challenges. Furthermore, if I make such an assertion, but don't evidence any of the attitudes or practical commitments that should attend it, then I've opened myself to reproach. The same assertion, if it is accepted, also gives us certain practical entitlements: to decide to φ, to punish someone who fails to φ, or to incorporate the judgment that one ought to φ into future moral reasoning. Understanding this socially enabled scorekeeping plays an essential role in establishing the conceptual role the moral 'ought' plays.

> It is this complex of what might be called "upstream and downstream inferential potential" that constitutes the inferential role of the statement. And, according to inferentialism, this role is basic in the explanation of the meaning of the statement. (Chrisman 2008, p. 350)

Again, this means that an inferential account of the meaning of the moral "ought" will need to establish both introduction and elimination rules for the use of the term—those rules that tell us what conditions warrant the use of the term, and what further inferences are licensed by its use, respectively. I'll give a suggestion of how such rules might look below, but it's important to keep in mind that the success or failure of metaethical inferentialism depends on the availability of such rules—and not necessarily on the plausibility of the particular sketch I'm offering here. Let's consider the elimination rules first.

U's assertion "one ought to φ" *licenses the following inferences*:[11]

---

[11] Again, the question of how this licensing comes about is outside the scope of this paper. It may come about as part of a socially directed practice of holding *U* responsible to certain kinds of commitments and entitlements, as per a normativist interpretation of inferentialism. Or, as per a naturalist interpretation of inferentialism (see Fn.2), it may be because sincere moral judgments necessarily involve some (possibly defeasible) motivational disposition. (In my formulation of the inferential rules for 'ought' here, I remain neutral on which interpretation is correct.).

R1: That U has attitudes in favor of φ-ing, including perhaps motivation to φ, a commitment to approve of one who φ's, and the second-order belief "one ought to be motivated to φ". These attitudes, commitments, and dispositions reinforce the practical clout of the moral 'ought'; their practical significance is not contingent on the desires or goals of particular agents.[12]

R2: That U has similar attitudes and commitments against not φ-ing, including perhaps a commitment to feel guilt upon failing to φ, disapproval or disgust towards not φ-ing, perhaps to the point of effecting punishment and even pressure on third parties to react punitively to those who fail to φ, and the second-order belief "one ought to disapprove of not φ-ing."

A complete inferential account will likewise specify the introduction rules that indicate the appropriate circumstances for use of the moral 'ought', language entry rules such as:

R3: If U has a disposition or commitment to approve of S's ψ-ing, to disapprove or feel disgust at S not ψ-ing, to punish S for not ψ-ing, to pressure third parties to react punitively to S's failure to ψ, etc., and if the practical significance of these commitments is not contingent on the desires or goals of particular agents, then U is *defeasibly* licensed to assert that S ought to ψ.[13]

Finally, the moral 'ought' can be used to license inferences to conclusions that involve the motivating features of moral discourse, but because we can use it with negations, we can also use 'ought' to *block* certain kinds of inferences. This gives us an intra-linguistic rule that shows the inferential relationship between our use of 'ought' and 'is permissible':

R4: $\sim O(\sim\varphi)\leftrightarrow P(\varphi)$: "One ought to steal" licenses the use of practical force in favor of stealing. It can be contradicted in two ways: "One ought to not steal"

---

[12] Cf. Wedgwood on the inferential import of the broadly deliberative 'ought', which he argues is always implicitly or explicitly indexed to some agent and time: "Acceptance of the first-person statement '$O_{<me,t>}(p)$'—where 't' refers to some time in the present or near future—commits one to making p part of one's plan about what to do at t." (2006, p. 137) Ignoring questions of indexing time and agent, we can see the similarity of his account to my own—if we keep in mind that mine is narrowly concerned with moral 'ought's, and don't flesh out the notion of a commitments and plans in psychological terms (i.e. of commitments and plans as kinds of mental states) but instead in terms of the kind of pragmatic influences (i.e. towards inferential commitments to arrive at deliberative conclusions about how to act) enumerated in Sect. 5 above. See also Charlow (2013, fn. 20).

[13] In Fn. 10 I considered the possibility that there are more substantive requirements on moral discourse; distinctive considerations about fairness, benefit, and so on may inform ethical claims, perhaps even to the point of rendering the acceptance of certain normative propositions as a precondition for the use of moral discourse. If there are further characteristic features of moral language—including normative propositions that have a status as conceptual truths—we should expect this to be reflected in the meaning-constituting inferential roles for the moral 'ought':

R5: If U recognizes that an act ψ is an instance of recreational slaughter/breaking a promise for convenience's sake/humiliating others simply for pleasure/etcetera, then all things being equal, U is defeasibly licensed to claim that one ought to not ψ.

These language entry rules will also have joint implications for language exit rules:

R6: U's assertion that one ought to ψ implies that ψ-ing is not an instance of recreational slaughter/promise-breaking for convenience's sake/humiliation others for pleasure/etcetera.

licenses practical force *against* stealing; "It's not the case that one ought to not steal", though, undermines such force—it gives *permission* to steal (more on this below).[14] This means then that U's assertion "one is permitted to φ" defeasibly licenses the inference that U is entitled to φ without feeling guilt, that U does not stand liable be motivated to φ, and does not stand ready to effect punishment or to pressure third parties to react punitively to those who fail to φ.

The kind of explanation I've offered here might raise a red flag. In arguing for an inferential understanding of the moral 'ought', I'm making use of the kinds of pro- and con-attitudes that are the hallmark of an expressivist account. One might suspect that I've slipped back into an expressivist mode of explanation. But the inferentialist doesn't have to deny that moral assertions express some sort of mental states—any more than she must deny that descriptive assertions express beliefs. What she will deny, in both cases, is that this expressive relationship determines the meanings of those assertions. To understand the inferential import of moral claims, we must look past the varied mental states that are generally correlated with sincere instances of moral claims, to the practical impact that these claims typically have.

> Perhaps sincere ethical claim-making is indeed... correlated with having the sorts of desires, intentions, and plans which explain motivation, but the exceptions to this psychological correlation indicate that it shouldn't dictate our explanation of the meaning of the relevant claims. We can recognize that certain sorts of ethical claims conventionally express desires, intentions, and/ or plans without holding that they mean what they do in virtue of expressing those things. (Chrisman 2010, p. 119)

The attitudes that attend these claims shift: sometimes the payoff of a moral statement is a direct motivation; sometimes it's guilt at a past transgression; sometimes it's fear of punishment. This flux of attitudes serves an underlying function—to promote or discourage particular types of behavior. And it is this function—not the attitudes that serve it—that explains the inferences that are licensed by utterances involving the moral 'ought'.

Following Williams' meta-theoretical framework, we're now in a position to explain the meaning of the moral 'ought' in terms of its use. The meaning-constituting inferential rules outlined above aren't bare stipulations formulated in order to offer a solution to the Frege-Geach problem; rather, they are justified by the way in which they support the pragmatic aim of moral discourse. We can see how a term with these inferential roles is able to serve the function of social-coordination.

Obviously, for moral discourse to constrain our behavior, we will need at least one term that influences our actions. Because the moral 'ought' carries behavior-influencing force with its particular practical clout, its effect on our behavior is stable—the import of an 'ought' claim is not contingent on any particular desires or

---

[14] Compare again Wedgwood's treatment: "Acceptance of the first-person statement '$P_{<me,t>}(p)$'—where 't' refers to some time in the present or near future—permits one to treat p part of one's plan about what to do at t." (2006, p. 137).

authoritarian commands. Because the term influences moral emotions, and because it can inform liability regarding punishment and reward, it can weigh against the selfish motivations of those whose behavior is not swayed directly by moral considerations. 'Ought' claims have implications for utterers' behavior, and the expectations others have regarding this behavior is essential for coordinating behavior. And because the inferential import of the term is defeasible, our moral claims and the motivational import that accompanies them are subject to revision and retraction, and so moral discussion affords interlocutors a space in which to reach consensus. Finally, for moral language to effectively facilitate coordinated social behavior, we need to know not only when to employ its typical motivational force, but also when to not employ it—introducing the notion of permissibility into moral discourse makes available a fine-grained application of its practical import that is otherwise out of reach.

7 Solving the Frege-Geach problem

To address the Frege-Geach problem, the inferentialist needs an account of the meaning of moral terms—an account that enables moral language to fulfill the functions described in Sect. 5. Consider the following argument:

(4) If he cheated, he ought to be punished.
(5) He cheated.
(6) He ought to be punished.

The 'ought' in the conclusion of this argument has a 'practical oomph': one's claim that he ought to be punished implies, for example, that one stands ready to punish him. But this 'oomph' is not present in its embedded form in the premise—asserting the conditional premise does not imply that one is in favor of punishment in the way that asserting the conclusion does. To escape charges of equivocation, the moral inferentialist has to explain the meaning of the embedded clause such that its content remains constant across these different assertions.

It's crucial here to maintain a distinction between the pragmatic import of moral discourse and the inferential roles that enable it to fulfill this function; we must avoid slipping "into thinking that use is at bottom *only* pragmatic significance, forgetting the use-patterns that fix conceptual content." (Williams 2010, p. 325) In making moral claims, we are not merely *doing* something but *saying* something. It is the meaning of the term and not what we do with it in moral utterances that remains constant even in embedded contexts. And according to Williams' framework, what we are saying—the content of moral claims—is explained by the inferential rules that fix the meaning of the moral term involved. As we've seen, these rules in turn are explained by the how they function to enable the pragmatic aim of moral discourse.

So according to the inferentialist, the meaning of the moral 'ought' is constituted by certain inferential rules that play a characteristic motivational role. If we accept that the term has these inferential roles whether it is embedded or unembedded, it

follows that the meaning remains constant and so we avoid the charge of equivocation associated with the Frege-Geach problem. Recall the analogy to the chess piece in Sect. 4: a pawn is in an important sense constituted by the rules of play that govern it. Just as we cannot understand a term like the moral 'ought' without understanding the inferential rules that inform its use, we cannot understand what a pawn is without understanding the rules of play that govern it. This does not mean that either the pawn or the term must display all the capacities articulated by the rules in any given move in the game, or in any given assertion in a conversation—sometimes a pawn will not or will even be unable to make any captures (as when it's in a position in which it is offering protection to other pieces), and sometimes the moral 'ought' does not directly guide our behavior (as when it's being used in a force-stripping context). But we would be unable to understand the state of play either in a conversation or in a game of chess without our grasp of the situation being informed by the relevant rules.

Consider the argument regarding the punishment of a cheater above. This is an instance of *modus ponens* that is the paradigmatic challenge of the Frege-Geach problem. Here, 'ought' is governed by the same inferential rules in both premise and conclusion: the meaning of the term remains constant. Our use of the term enables us to reason from a principle calling for behavior-influencing force in a particular situation that, combined with an understanding that one is actually in that situation, licenses an inference to a conclusion that brings the relevant practical force into play. One can see why it makes sense for us to use a claim involving 'ought' to reason with conditionals. To effect complex social coordination with moral language, the motivational force associated with moral assertions should be situation-specific; conditionals with descriptive antecedents and normative consequences enable one to fulfill this function. And with some ingenuity, the reader might imagine how normative reasoning with logical connectives like **&** and ∨ also enable the pragmatic role of moral discourse to be fulfilled.

There's a very important distinction to note here in the explanation an inferentialist can give of the validity of an argument like the above, and the kind of explanation required of an expressivist. This is because the expressivist's "hypothesis that some sentences express mental states that are not beliefs sets a *constraint* on the accounts of each kind of sentential connective that is not faced by non-expressivist semantics." (Schroeder 2008a, p. 178). I'll explore this constraint in greater depth in Sect. 10, but the idea is this: For any logically complex claim involving moral terms (like the premise in the argument above), the expressivist needs to give an account of what sort of *mental state such a claim expresses.* She must then show why it is that such a state of mind provides the sentence in question with the semantic properties we'd expect—and why it is that such a sentence has the same kind of inferential import as it would if it were straightforwardly descriptive. So for the argument above, the expressivist takes on a burden of providing an adequate semantic account of the meaning of the conditional premise and why it licenses *modus ponens.* But the burden isn't limited to just that. As Mark Schroeder argues, an adequate expressivist account of the semantics of moral language has to explain why it is that.

moral terms have a different kind of semantics than ordinary descriptive terms, but somehow every complex-sentence-forming construction manages to do exactly the same sort of things with them that it does with ordinary descriptive terms. (Schroeder 2008b, p. 714)

This means that to fully discharge the burden, the expressivist must give an account of the mental states expressed by any logically complex normative statements—such an explanation will have to cover the semantic properties not only for conditionals, but also for negations, conjunctions, disjunctions, quantifiers, modals, tense, and so on.[15]

This is a very heavy burden, and one of the principle advantages of inferentialism is that this is a burden the inferentialist doesn't have to shoulder. From this perspective, the use of the logical connectives can come for free: unlike the expressivist, the metaethical inferentialist doesn't have to give a piecemeal explanation of the semantics for all the different contexts in which moral claims are used. Now that we've established language-exit and language–entry rules for 'ought' and the language–language rules that relate 'ought' to 'is permissible that', and note that these can be placed outside of any well-formed sentence φ, such that $O(\varphi)$ and $P(\varphi)$ qualify as assertions, we have a *prima facie* case that these terms act as sentential operators, and as such we can use them with standard sentential connectives to make complicated moral sentences. The validity of arguments involving those connectives would then follow from the meaning of the connectives themselves. As we will see in Sect. 9 below, metaethical inferentialist can make use here of an inferentialist account of logic, and explain the validity of any inference as consequences of the introduction and elimination rules that come with the logical connectives.[16] From this perspective, logic is understood as topic neutral; *modus ponens* is logically valid, regardless of the subject matter.

Nor should one worry that we can't get an understanding of validity without an understanding of representationalist-style truth conditions, for the inferentialist will of course reject this understanding of validity. The inferentialist account I'm considering here begins at the outset with the notion of correct inferences—those that a competent language-user will be expected to have mastered in virtue of understanding the meanings of the premises and conclusion—and explains valid arguments as ones in which one is entitled to the conclusion simply in virtue of being committed to the premises.[17] In the case of the moral 'ought', the network of

---

[15] In *Being For*, Schroeder attempts this Herculean task on behalf of the expressivist; his account involves so many complications and problematic assumptions that he concludes it might just amount to a *reductio* of the expressivist project. (Schroeder 2008a, pp. 92, 177) See Marbito (2009), Alwood (2010), and Wedgwood (2010) for critical responses.

[16] See (Brandom 1994), (Beall and Restall 2013), and (Ripley 2013) for inferentialist approaches to validity.

[17] For those who insist that an adequate account of validity must involve truth-preservation, the inferentialist can be confrontational or conciliatory. Confrontationally: There is good reason to think that the relationships of consistency and inconsistency that are fundamental to an explanation of validity needn't be underwritten by a truth-theoretic account; following Portner (2010) and using a meaning-from-use approach not dissimilar to my own, Charlow (2013) argues persuasively that we can (and should) extend a non-representational account of (in)consistency between imperative sentences to declarative

commitments and entitlements that make sense of the validity of moral arguments aren't mere stipulations; they are reflected in the inferential rules that we'd expect the term to have, given its role in promoting social coordination.

The sketch above fits the desiderata I listed at the beginning of the last section for an adequate response to the challenge of the Frege-Geach problem:

C1. It does not presuppose any domain of elusive moral facts that are waiting to be described or responded to. Instead it makes use of a naturalistically respectable understanding of the behavior-guiding function we might expect organisms like ourselves to use our language to fulfill.

C2. The inferential rules licensed by the conceptual role played by 'ought' gives us the meaning of the term in such a way that it is clear this meaning remains constant even in force-stripped contexts, so that the charge of equivocation is evaded.

C3. Identifying the inferential role of moral terms helps us understand the validity of inferences that we make involving the moral 'ought'. These inferences are not understood as valid simply because of the representation-style truth conditions of the premises and conclusions. Instead the inferences are licensed by the use we make of moral language.

Having outlined how an inferentialist account can solve the Frege-Geach problem, I want to argue that it also gives us the tools to deal with at least some of the worrying criticisms that have been leveled at the standard expressivist solution. Anyone familiar with contemporary literature on expressivism knows these are numerous; I'm focusing on two that have proven particularly malignant: the problem of permissions, and the expressivist commitment to what Schroeder calls 'mentalism'. I'll explain each, and show how an inferentialist account can either avoid the problem altogether, or at the very least render it more manageable.

8 The problem of permissions

The standard expressivist response to the Frege-Geach problem involves understanding the inferences in terms of the logical relationships that inhere between the mental states expressed by the different parts of the inference. So for example, two sentences contradict one another when it would be inconsistent for one to be in both of the mental states these sentences express. Consider two moral claims that are obviously contradictory:

(7) You ought to vote for Obama.
(8) You ought to not vote for Obama.

Footnote 17 continued

normative claims. Conciliatorily: This shouldn't be taken as an outright denial that validity (in moral arguments like the above, at least) entails truth-preservation, though: inasmuch as the inferentialist makes use of a minimal approach to truth, she can grant that valid arguments with true premises are guaranteed to have true conclusions.

(7) expresses some mental attitude towards voting for Obama—let's say approval, to keep things simple. And (8) represents the same attitude towards *not* voting for Obama. The two claims contradict one another, on the expressivist account, because one cannot consistently approve of *p* and approve of *not p*; one could not consistently have the *same sort of attitude* toward *contradictory content*.

Such an explanation has an advantage: the notion of contradiction between mental states here is one that is not specific to expressivism. We can think of other instances where commonsense dictates two mental states contradict one another *because they are the same sort of attitude toward contradictory content*—a belief that *p* contradicts a belief that *not p*, an intention to *q* contradicts an intention to *not q*, and so on.

Nicholas Unwin (1999, 2001) points out there is more than one way to contradict (7); how are we to make sense of the contradiction between (7) and the following?

(9) It's not the case that you ought to vote for Obama.

Obviously, the expressivist must account for the contradiction between (7) and (8) in terms of the mental states the two claims express, but (8) does not express the same mental state as (9). And of course, we cannot simply explain the mental state (9) expresses as a kind of refusal to accept (7), since, as Unwin points out, "[T]here is certainly a very real difference between not accepting (or refusing to accept) something and actually accepting its negation." (Unwin 1999, p. 341) It's very tempting to respond that (9) doesn't express the same kind of mental state as (7) and (8)—two mental states of the same kind with contradictory content—but instead expresses a different type of mental state altogether: *tolerance*. The idea is that the contradiction between (7) and (9) doesn't come from having the same kind of attitude towards contradictory content, but from having *different kinds of attitudes* to the *same content*.

The problem is that such an approach leaves unexplained *why* different attitudes like tolerance and disapproval would be related to one another logically. Common sense gives us a model of logical inconsistency between type-identical mental states that have inconsistent content, but we'd need a new model to explain mental states that are inconsistent because they involve different attitudes towards the same content. The expressivist could of course just stipulate that claims expressing attitudes of disapproval and tolerance have the inferential relationships they do, but making such a decree doesn't do any work to explain why those relationships inhere, or what sort of logical relationships we should expect them to have. "It becomes at best a mere brute fact that the attitudes conflict with each other, with no internal complexity that could explain why." (Ibid., p. 342) The worry is that "psychological states are not endowed with the right sorts of properties to offer self-sufficient explanations"(Charlow 2013, p. 20) of why these states are inconsistent.

At the beginning of my account of inferentialism, I argued that one of its virtues is its simplicity; all things being equal, one should prefer a semantic account that directly explains those inferential roles that are constitutive of the meaning of a term over an account that explains meaning by reference to mental states which themselves provide the appropriate inferential roles. Now we can see that such a

virtue might pay dividends: because the inferentialist does not rely directly on (in)consistency between mental states to account for inferential relationships, she doesn't need to posit tolerance as a special sort of mental state with its own unique permission-giving inferential features. She can instead account for the inferential role of 'permission' itself to explain its fundamental inferential features. Here the inferential rule R4 outlined above is fulfilled. To permit behavior is (among other things) to withhold the practical forces associated with 'ought'; with negation this can be expressed by saying that "is it not the case that you ought not p", or simply "p is permissible".

With this account of the inferential role the term plays in hand, we can understand the different kinds of negation that have vexed expressivist accounts. (7) and (8) above contradict one another because in uttering (7), one stands committed to invoke behavior-influencing force in favor of voting for Obama, whereas (8) invokes the same force *against* voting for Obama. The contradiction between these claims "consists in their potential, through inference, to engender conflicting desires and decisions." (Horwich 2010, p. 183)[18] Likewise, (7) and (9) contradict one another because (9) undermines the very force that (7) invokes. Where (8) invokes the characteristic behavior-influencing force of 'ought' against voting for Obama, (9) invokes another sort of behavior-influencing force, which can at least in part be understood in terms of the kind of typical reaction someone voicing such a permission is liable to have to someone who says we shouldn't vote for Obama. More generally, if U asserts "it's permissible to φ", this implies (among other things) that U lacks dispositions to feel guilt upon φ-ing, to disapprove, punish, or ostracize someone who has φ-ed—and furthermore this implies U has the second-order commitment that one ought to not disapprove of φ-ing, where such a second-order belief would have its own dispositional consequences (e.g., to chide those who say things like "one ought to not φ"). Again, the notion of negation here relies on how the two claims reflect commitments and entitlements that are mutually exclusive.

This explanation is not a bare stipulation about the inferential role of "is permissible", and it's not an *ad hoc* response to the problem. Rather, it follows naturally from Williams' approach to explaining meaning in terms of the inferential rules that enable the pragmatic function of a discourse. We use moral language to coordinate effective social behavior with terms that influence behavior with motivating force. As we saw in Sect. 6 above, the inferential rules governing 'is permissible' gives us fine-tuned control of this force—it allows us to indicate where such force is to be used, and where it is to be withheld. Because she grounds the semantic content of moral claims in the inferential role played by the terms themselves—rather than in the mental states the terms serve to express—the inferentialist doesn't have to insist that tolerance is a mental state with unique inferential properties, and so doesn't have to account for such properties.

---

[18] For more on cashing out negation in terms of pragmatic conflict, see Dreier (2009, especially pp. 103–106). In Warren (2013, pp. 128–168), I argue that an inferentialist account of this kind of conflict enjoys advantages over an expressivist account, according to which such conflict must be always be mediated by the relevant conative attitudes.

9 The commitment to 'mentalism'

Another problem for expressivism is the worrying possibility that in order to give a consistent treatment of sentential connectives, the expressivist must treat all such connectives—whether they stand between descriptive or normative atomic sentences—as expressions of mental states. Consider the following complex claims:

(10) If stealing is wrong, so is murder.
(11) If stealing is wrong, then Todd's in jail.
(12) If he stole the diamond, then we ought to tell the police.
(13) If he stole the diamond, then Todd's in jail.

Expressivists account for the meaning of (10) by holding that the indicative conditional expresses a mental state whose content is a function of the attitudes that would be expressed by its constituent parts—disapproval of stealing and disapproval of murder. What are we to make of sentences like (13), whose atomic parts don't both express normative states? And what about (11) and (12), whose atomic parts vary? One option is to hold that in each of the sentences above, the meaning of the conditional is different, and to introduce new accounts for each of the logical connectives: the meaning of the conditional in (10) would be given as a function of the mental states that would be expressed by its antecedent and consequent; the meaning of the conditional in (13) would be given, presumably, by a truth function between the propositions referred to in its antecedent and consequent; the meanings of the conditionals in (11) and (12) would each be given as a function between the mental state expressed in its antecedent or its consequent, respectively, and the proposition referred to by the complement.

Such an approach is not just messy but rather implausible, first because our use of all of the sentential connectives—conditionals, conjunctions, disjunctions, and negations—in *natural* language is not sensitive to the normative/descriptive distinction; we don't indicate our conditionals are functioning differently when they govern normative versus descriptive sentences. Secondly, an account of the sentential connectives that is not univocal will have difficulty explaining a plausible feature of logic:

> The inconsistency between 'P' and '∼P' should be guaranteed by the meaning of '∼', not by the joint fact that if we first interpret 'P' as normative and '∼' as normative-sentence-negation, they are inconsistent, together with the fact that if we interpret 'P' as descriptive and '∼' as descriptive- sentence-negation, they are inconsistent, and if we interpret it in any other way it is not well formed. (Schroeder 2008a, p. 22)

So there is considerable pressure to admit that the conditional plays the *same* fundamentally expressive role in all of these conditionals. "If any of the conditionals [in (11)–(13)] are the same as the conditional [in (10)], then it must be the *same function* from two mental states to a third mental state." (Schroeder 2008a, p. 95, emphasis mine) This entails what Schroeder calls 'mentalism', the view that *descriptive claims* get their meaning (and semantic properties) from the mental states—presumably the *beliefs*—they express. This holds for complex

descriptive sentences like (13) above, but also for its atomic parts. The meaning of conditionals—or any logical connectives governing atomic sentences—is given as a function of the mental states expressed by their parts, so it follows that the unembedded atomic clauses of (13) must express mental states in the same way as the complex sentence does.

Mentalism is, Schroeder puts it mildly, "non-trivial" (Ibid., p. 23). The expressivist approach to the semantics of complex normative claims forces us into a substantive claim about the content of descriptive claims. On the "standard" semantic picture, the content of a sentence—and the belief that it expresses—is given by the truth-conditions for the proposition to which it corresponds. But on the expressive semantic picture, this is turned on its head: the content of a descriptive sentence doesn't come from the proposition to which it corresponds, but is instead derived from the mental state that it expresses (Ibid., pp. 31–34).

None of this is conclusive. Mentalism may, after all, be the right account of descriptive content. It does show us, however, that the expressive approach to semantics—in which we understand the semantic properties of normative sentences in terms of the mental states they express—cannot be made in a vacuum. It has entirely novel implications for fundamental semantic issues.

Inferentialism distinguishes itself from expressivism on precisely those grounds that lead to the troubling charge of mentalism, so one might suppose that the charge has no weight here. After all, if the inferentialist doesn't account for the meaning of moral claims in terms of the mental states they express, there's no reason to think that inferentialism needs to account for the meaning of logical connectives or descriptive claims in the same way. But even though the specifics of Schroeder's charge get no traction in regards to inferentialism, we may be worried that an argument of the same general form applies.

On the inferentialist account, the meaning of the conditional in (10) would be given as a function of (not the mental states but rather) the inferential roles played by its antecedent and consequent. For reasons parallel to those Schroeder lays out, the meaning of the other conditionals must likewise be given as a function of the inferential roles played by its atomic parts—whether normative or descriptive. It follows then that the meaning of all atomic claims that can be combined by logical connectives with moral claims are to be given by their inferential roles! Though such a view wouldn't rightly be called 'mentalism' it is nonetheless a non-trivial position about fundamental semantic issues.

The conclusion is compelling, but not nearly so troubling as the charge of mentalism. It may be non-trivial, but as we've seen it certainly isn't entirely novel. There already is a plausible theory of linguistic meaning committed to such a conclusion: inferential role semantics, according to which the meaning of expressions can be explicated by the inferences they license and by which they are licensed. Perhaps it may come as a surprise that metaethical inferentialism entails a general inferentialist account of meaning for non-normative claims, but even if it does, the account finds itself in good company. Indeed, as we saw in Sect. 7, the access a metaethical inferentialist enjoys to the broad semantic resources of a general inferential approach counts as a considerable advantage over expressivism. Accommodating an inferentialist approach to logic—an independently motivated approach

that is tailored to handle questions of validity and systematic compositionality—makes resources available to the metaethical inferentialist to explain why moral terms can figure in all of the same sorts of arguments as straightforwardly descriptive ones, without appealing to representational correspondence with truth makers in their explanation. For the inferentialist, Schroeder's semantic challenge may not be entirely undermined, but it is defanged. This is in stark contrast to the challenge expressivism faces, which requires a novel (and some may charge, *ad hoc* and radical) understanding of how mental states inform the meaning of our claims.

10 Where this leaves us

It has been my intention with this paper to sketch a new way to respond to an old problem, in the hopes that this will motivate a more complete project of giving an account of the semantics of moral language in inferential terms without depending on the expression of mental states. I've very briefly sketched two of the problems that expressivism inherits because of its dependence on such states, and inferentialism's advantages in dealing with these. A point-by-point reckoning with all of the semantic issues plaguing expressivism lies outside the scope of this paper, but my hope is that I've made the case that metaethical inferentialism merits more attention.

One would obviously also need an account of the myriad other moral concepts we use, concepts like 'good' and 'bad', but also thick moral concepts—those which include both evaluative and descriptive elements—like 'brave' and 'miserly'. Astute readers might also notice that my treatment of 'ought' is incomplete, for two reasons. First, 'ought' isn't always used morally; presumably a complete account will integrate our moral and non-moral uses of the term. Second, we often use the moral 'ought' in contexts where the behavior-influencing force of moral language doesn't obviously have a place. What sort of behavior are we trying to effect when we say, "Stalin oughtn't have been so devious"? And what are we to make of moral claims that don't directly involve any agents at all, as in, "This shopping mall ought to be more accessible to the disabled"?[19] Speculatively, it seems the inferentialist has two options: Either seek a reduction of such claims in terms dissectible via the pragmatic function of moral discourse I listed above, or expand on the list to account for such uses. Assessing the relative plausibility of these options is a project for another time.

And of course we can't forget those other semantic problems expressivism faces that I referenced earlier but haven't addressed: G.F. Schueler argues that non-cognitive mental states like approval and disapproval just aren't plausible candidates as manufacturers of genuine logical contradiction. Mark van Roojen points out that the expressivist account of validity in an argument seems to

---

[19] For Wedgwood's suggestions on giving a more general treatment of 'ought', see (2006, pp. 151–157). As I've explained in Section 6, my account of 'ought' differs from his in important ways; nevertheless, I see no reason that his account of the various implications of 'ought' couldn't be favorably extended to a non-representational approach such as my own.

overproduce valid arguments; it gives us a model that attributes validity to arguments that are obviously invalid. David Merli argues that expressivism faces serious difficulties giving an account of the mental states being expressed in moral disagreements in such a way that explains why it is that they actually count as disagreements. Perhaps some of these issues will dissolve when we take up the inferentialist perspective; all of them are critiques aimed at expressivism, after all, and involve problems that spring from the mental-state-expressed account of content that inferentialism rejects, so one might hope they simply don't apply (this seems especially promising regarding Schueler's objection). Still, one could be worried that some or all of these criticisms can be leveled in an inferentialist key— indeed, that they might arise for *any* non-representational account of content. Responding to all of these concerns on behalf of inferentialism is another deep project.

I think these projects hold real promise, but of course the devil is in the details. Much work needs to be done to convincingly show that the approach resolves these issues. Because inferentialism offers us a simpler view of moral semantics, and because it promises to skirt many of the semantic issues that plague expressivism, I believe these are projects that are worth our attention.

References

Alwood, A. (2010). Imperative clauses and the Frege-Geach problem. *Analysis, 70*, 105–117.

Ayer, A. J. (1936/1946). Language, Truth and Logic, London: V. Gollancz ltd.

Bar-On, D., & Chrisman, M. (2009). Ethical neo-expressivism. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 4, pp. 133–165). Oxford: Oxford University Press.

Beall, J. C. (2009). *Spandrels of truth*. Oxford: Oxford University Press.

Beall, J. C., & Restall, G. (2013). Logical consequence. In: Edward N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Winter 2013 Edition) http://plato.stanford.edu/archives/win2013/entries/logical-consequence/.

Blackburn, S. (1984). *Spreading the word: Groundings in the philosophy of language*. New York, USA: Oxford University Press.

Blackburn, S. (1988). Attitudes and contents. *Ethics, 98*(3), 501–517.

Blackburn, S. (1998). Attitudes and contents. *Ethics, 98*, 501–517.

Blackburn, S. (2006). The semantics of non-factualism. In M. Devitt & R. Hanley (Eds.), *The Blackwell guide to the philosophy of language*. Oxford: Blackwell.

Block, N. (1986). Advertisement for a Semantics for Psychology. *Midwest Studies in Philosophy, 10*, 615–678.

Boyd, R. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism*. Ithaca: Cornell University Press.

Brandom, R. (1983). Asserting. *Noûs, 17*(4), 637–650.

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Cambridge, MA: Harvard University Press.

Brandom, R. (2007). Lecture four: Modality and normativy: From Hume and Quine to Kant and sellars. *John Locke Lectures*. Lecture conducted from Oxford University, Oxford.

Brandt, R. (1979). *A theory of the good and the right*. New York: Oxford University Press.

Brink, D. O. (1989). *Moral realism and the foundations of ethics (Cambridge Studies in Philosophy)*. New York: Cambridge University Press.

Charlow, N. (2013). The problem with the Frege-Geach problem, *Philosophical Studies,* 1–31.

Chrisman, M. (2008). Expressivism, inferentialism, and saving the debate. *Philosophy and Phenomenological Research, 2,* 334–358.

Chrisman, M. (2010). Expressivism, inferentialism, and the theory of meaning. In M. Brady (Ed.), *New waves in metaethics*. Basingstoke: Palgrave-Macmillan.

Cuneo, T., & Shafer-Landau, R. (2014). The moral fixed points: New directions for moral nonnaturalism. *Philosophical Studies*, Manuscript in preparation.

Dreier, J. (2006). Negation for expressivists: A collection of problems with a suggestion for their solution. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 1, pp. 217–233). Oxford: Oxford University Press.

Dreier, J. (2009). Relativism (and Expressivism) and the problem of disagreement. *Philosophical Perspectives, 23*(1), 79–110.

Field, H. (1977). Logic, meaning and conceptual role. *Journal of Philosophy, 69,* 379–409.

Geach, P. T. (1965). Assertion. *The Philosophical Review, 74,* 449–465.

Gibbard, A. (1990). *Wise choices, apt feelings: A theory of normative judgment*. Cambridge, MA: Harvard University Press.

Gibbard, A. (1994). Meaning and normativity. *Philosophical Issues, 5,* 95–115.

Gibbard, A. (2003). *Thinking how to live*. Cambridge, Mass: Harvard University Press.

Hare, R. M. (1952). *The language of morals*. Oxford: Oxford Clarendon Press.

Harman, G. (1999). *Reasoning, meaning and mind*. Oxford: Oxford University Press.

Horgan, T., & Timmons, M. (2006). Cognitive expressivism. In T. Horgan & M. Timmons (Eds.), *Metaethics after moore* (pp. 255–298). Oxford: Oxford University Press.

Horwich, P. (1998). *Meaning*. New York: Oxford University Press.

Horwich, P. (2010). *Truth-meaning-reality*. New York: Oxford University Press.

Jackson, F., & Pettit, P. (1995). Moral functionalism and moral motivation. *Philosophical Quarterly, 45,* 20–40.

Joyce, R. (2006). *The evolution of morality*. Cambridge, Massachusetts: MIT Press.

Loar, B. (1981). *Mind and meaning*. Cambridge: Cambridge University Press.

Mabrito, R. (2009). *Notre Dame Philosphical Reviews, Review of Mark Schroeder's Being For*. Retrieved October 18, 2012. http://ndpr.nd.edu/news/24142-being-for-evaluating-the-semantic-program-of-expressivism/.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. London: Penguin.

Merli, D. (2008). Expressivism and the limits of moral disagreement. *The Journal of Ethics, 12,* 25–55.

Peacocke, C. (1992). *A study of concepts*. Cambridge, Massachusetts: MIT Press.

Portner, P. (2010). Permission and choice. Manuscript, Available from semanticsarchive.net.

Price, H., & Macarthur, D. (2007). Pragmatism, quasirealism and the global challenge. In C. Misak (Ed.), *The new pragmatists*. Oxford: Oxford University Press.

Prinz, J. (2007). *The emotional construction of morals*. Oxford: Oxford University Press.

Ripley, D. (2013). Paradox and failures of cut. *Australasian Journal of Philosophy, 91*(1), 139–164.

Ryle, G. (1949). *The concept of mind*, Hutchinson and Co.

Schroeder, M. (2008a). *Being For*. Oxford: Oxford University Press.

Schroeder, M. (2008b). What is the Frege-Geach problem? *Philosophy Compass, 3*(4), 703–720.

Schueler, G. F. (1988). Modus ponens and moral realism. *Ethics, 98,* 492–500.

Smith, M. (1994). *The moral problem*. Oxford: Blackwell.

Sturgeon, N. (1988). Moral explanations. *Sayre-McCord, 1988,* 229–255.

Thomasson, A. (2007). *Ordinary objects*. New York: Oxford University Press.

Thomasson, A. (2014). *Ontology made easy*. New York: Oxford University Press.

Timmons, M. (1999). *Morality without foundations: A defense of ethical contextualism*. New York: Oxford University Press.

Unwin, N. (1999). Quasi-realism, negation and the Frege-Geach problem. *The Philosophical Quarterly, 49,* 337–352.

Unwin, N. (2001). Norms and negation: A problem for Gibbard's logic. *The Philosophical Quarterly, 51,* 60–75.

van Roojen, M. (1996). Expressivism and irrationality. *Philosophical Review, 105,* 311–335.

Warren, M. D. (2013). Lightweight moral realism: Objectivity and reasoning without heavyweight facts. *Open Access Dissertations.* Paper 1138. http://scholarlyrepository.miami.edu/oa_dissertations/1138.

Wedgwood, R. (2006). *'The meaning of "Ought"', Oxford studies in metaethics* (Vol. 1, pp. 127–160). Oxford: Oxford University Press.

Wedgwood, R. (2010). Schroeder on expressivism: For—or against? *Analysis, 70,* 117–129.

Williams, M. (2010). Pragmatism, minimalism, expressivism. *International Journal of Philosophical Studies, 18,* 317–330.

Williamson, T. (2003). Understanding and inference. *Aristotelian Society Supplementary Volume, 77*(1), 249–293.

Williamson, T. (2007). *The philosophy of philosophy*. London: Blackwell

Wong, D. (1984). *Moral relativity*. Berkeley: University of California Press.