Handling Rejection

Abstract

This paper has two related goals. First, we develop an expressivist account of negation which, in the spirit of Alan Gibbard, treats disagreement as semantically primitive. Our second goal is to make progress toward a unified expressivist treatment of modality. Metaethical expressivists must be expressivists about deontic modal claims. But then metaethical expressivists must either extend their expressivism to include epistemic and alethic modals, or else accept a semantics for modal expressions that is radically disjunctive. We propose that expressivists look to Amie Thomasson's work for a general strategy for offering a unified expressivist account of modality. Modals in general, we propose, are devices for expressing metalinguistic commitments within the object language, with deontic, epistemic, and metaphysical modals all expressing different kinds of metalinguistic commitments.

1 Introduction

This paper has two related goals. First, we develop an expressivist account of negation which, in the spirit of Gibbard (2003), treats disagreement as semantically primitive. We show that this sort of account can be constructive and extensionally adequate. Of course, to say that disagreement is primitive in the semantic theory is not to say that it is primitive full stop. An explanation of why sentences disagree will be the work of a metasemantic theory, and we will sketch one such explanation below. By providing both an adequate expressivists semantics of 'not' and an explanation of why 'P' and 'not P' disagree, we will have a solution to the Negation Problem for expressivists (Unwin 1999 and 2001; Dreier 2006; and Schroeder 2008).

Our second goal is to make progress toward a unified expressivist treatment of modality. In our opinion, metaethical expressivists are already implicitly committed to such an account. As metaethical expressivists, they must be expressivsts about deontic modal claims. But then they must either extend their expressivism to include epistemic and alethic modals, or else accept a semantics for modal expressions that is radically disjunctive.

We think expressivists should look to Amie Thomasson's (2007, 2013, 2017 and 2020) work for a general strategy for offering a unified expressivist account of modality.[1] For Thomasson, expressions such as 'it's necessary that', at least in metaphysical contexts, should be understood in a deflationary, inferentialist sense. The meaning of the expression is given, not by reference to some set of possibilities, but by a pair of inference rules governing when one is justified in concluding that something is necessary, and what sorts of conclusions follow (2020: 83-84). In virtue of this inferential role, speakers can then use claims of metaphysical necessity or possibility to make explicit, endorse, or negotiate about the semantic rules governing discourse, while remaining within the object language (2020: 64ff.).

Our position can be understood as a generalization on this position, adapted slightly for an expressivist rather than inferentialist framework: modals are a device for expressing *metalinguistic commitments* within the object language. Metaphysical modals express analytic commitments (e.g., definitional or semantic commitments), whereas epistemic modals express commitments regarding the *assertability* of claims, and deontic modals express commitments regarding the *prescribability* of imperatives. 'It might be raining' expresses a commitment to object to any assertion inconsistent with it being raining. 'Bob may stand up' expresses a commitment to object to any command to Bob to stay seated.[2] We will develop in this paper a constructive–if admittedly simplified and idealized–account of epistemic and deontic modality. Our engagement with metaphysical modality will be limited at this point to gesturing at Thomasson's work, but the larger goal is to provide a constructive account that unifies all uses of modal operators.

These commitments themselves should not be understood as simply another mental state (cf. Blackburn 1988) but rather as social commitments to one another, along the lines of (Woods 2018a and -b). By means of certain assertions, we put ourselves on the hook to one another for being or not being a certain way. In the case of something like an assertion of 'It might be raining', we make ourselves answerable to criticism and such by not following through on objecting to assertions of others that are inconsistent with it raining.[3]

The account is thus something of a hybrid between more traditional expressivist accounts and related inferentialist accounts such as those of Brandom (1994), Chrisman (2016), Thomasson (2020) and Tienfesee (forthcoming). Our account differs from inferentialism in that we do not take ordinary assertions to express public commitments to infer or public commitments to back up our mental

[1] We generally think unification a theoretical virtue, especially when the case seems to cry out for it, as the present one does.

[2] We agree that 'may' in English is ambiguous between a deontic and an epistemic reading. We will pretend this is not the case for purposes of exposition, but the ambiguity does not create any substantive problems for the view so far as we can tell.

[3] Of course, this story itself is complicated and beyond the scope of this paper, see (Woods 2018a and -b) for some details of how it might go.

states and actions with appropriate justification. Rather, public commitments to instantiate certain mental states, justified or no. This is what makes our view expressivist. Only for the devices we say express metalinguistic commitments—deontic modals, alethic modals, and logical operators—is there any non-trivial connection to instantiating a pattern of mental states, which is, in turn, the only part of our story which could be connected with inference itself. On inference itself, we follow Harman (1986) in thinking that.it is an action and one to be strongly distinguished from *implication* or *logical consequence*, which is a relation between (interpreted) sentences.

Our proposal also draws on ideas advanced elsewhere, especially on the work of Andrew Alwood (2016), who proposes treating permissibility claims as devices for 'blocking' *imperatives* from being added to a conversational scoreboard; the work of Huw Price (1990) on the origins and function of negation in terms of a primitive notion of incompatability,[4] the work of Matthew Chrisman (2016) suggesting that modal claims are unified by a similar 'metaconceptual' role, and the work of Luca Incurvati and Julian Schlöder (2017, 2019 and *forthcoming*), who propose treating 'might'-claims as devices for blocking *assertions* from being added to a conversational scoreboard. The proposal here is novel, in the first place, in that it provides a single integrated account of how both epistemic and deontic possibility claims serve to 'block' other speech acts, and secondly, in its attempt to give a more detailed account of the metasemantics of disagreement, of *why* various claims disagree.

We start by explaining what we mean by taking disagreement to be semantically primitive.

2 Disagreement and Meaning

What does it mean to take disagreement as semantically primitive, and why should we do so? First, it means treating certain disagreement relations between sentences as basic components of their meaning. We do not further explain disagreement in terms of reference. For contrast, consider how a more orthodox referential semantic theory would explain the disagreement between these two sentences:

(a) The sun is a star.
(b) The sun is not a star.

[4]Price however treats negation, roughly, as simultaneously functioning as a modifier of content and a modifier of force; for us, it only plays one of these roles at a time. He does this to avoid Fregean arguments against denial; we see no reason to address these, as they've been adequately addressed elsewhere, such as in Restall (2005).

The disagreement is in virtue of 'not' occurring in (b). Given this 'not', (b) denotes the complement of the set of worlds denoted by (a). The sentences then disagree because their intersection is empty. Alternatively we might explain the meaning of the 'not' in (b) truth-functionally, so (b) is true just in case (a) is false; and so the sentences disagree because they cannot both be true.[5]

On our view, 'not' is rather a way of lexicalizing disagreement. The meaning of (b) could be given simply as *disagreement with whatever (a) expresses*[6]. To foreshadow our more complete account, (a) and (b) should be understood as referring to the same proposition p (in simple embeddings), but (a) expresses *acceptance* of p whereas (b) expresses *rejection* of p. Disagreement is not to be understood in terms of some sort of incompatibility of content or other semantic objects, but adopting incompatible stances to one and the same content (Price 1990; Rumfitt 2000; Restall 2005; and Incurvati and Schlöder 2019).

To such "inferentialist" views it is sometimes objected that we can't understand what it is to reject a claim except in terms of accepting the claim's negation. Classical referentialist accounts of negation accept this point, reductively analyzing rejecting a claim in terms of accepting its negation. Yet the question at issue is precisely whether the use of logical vocabulary forces a referentialist interpretation of some domain of discourse.[7] We are suggesting an alternate interpretation: asserting (a) can be understood as expressing a commitment regarding the proposition that *the sun is a star*. Asserting (b) can be understood as expressing a commitment that rules the first commitment out.

Before saying more about the nature of these commitments, it is worth addressing a few immediate objections, and spelling out some of the motivation. To say rejection *rules out* acceptance is to say that by rejecting a claim one is committed to *not* accepting the claim. One may thus more seriously worry that we are defining rejection in terms of negation. We are, but there are no noncircular definitions of basic logical terms, as they are so basic to meaning and thought (Dummett 1974).[8] Truth-functional and set-theoretic definitions of negation likewise use the word 'not' in their metalanguage. The question is not whether analyzing negation in terms of rejection is reductive, but whether it is *constructive* in the sense of explaining how negation works.

We should also note that taking disagreement as primitive within a semantic theory is not the same thing as taking it as primitive *full stop*. We do not

[5] These explanations are obviously incomplete sketches. Also see Price (1990) for criticisms of this kind of explanation. Most notably, Price thinks this explanation makes it obscure why languages need expressions of denial. He argues that the role of negation in facilitating disagreement provides a clearer explanation of why we need an expression of negation.

[6] It might help some to think of the German particle 'doch' in this context. It works in a similar way, functioning as an explicit expression of disagreement with a previous statement.

[7] Similar points are made in (Incurvati and Schlöder 2019: 8).

[8] Of course, we can sometimes define one logical particle without recourse to itself by use of another. The natural analog of that point is also true here.

take disagreement to be primitive full stop, a position we regard as extremely unattractive. We assume that objects with semantic properties have those properties in virtue of something. But such an explanation is outside of the semantic theory proper: it is a part of metasemantics.

An analogy will be helpful here. A more orthodox semantic theory will take certain referential relations to be primitive, but this does not mean that it is committed to primitivism about reference. It is simply that for the purposes of offering a semantic theory, we are entitled to assume that terms refer. Why they refer and why they refer to the things they do are problems for metasemantics. Similarly, an expressivist is entitled to posit stances that stand in relations of disagreement. How exactly to fully explain disagreement should be regarded as a metasemantic problem for the expressivist. We sketch a metasemantics of disagreement in **section 4**. But note that while the semantic theory and the metasemantic account of disagreement are meant to work together, they are in principle independent: one could take on the semantic story but insist on a different explanation of why sentences disagree.[9]

Why treat disagreement as semantically primitive? For one thing, theorists like Gibbard have used this approach to tell attractive expressivist stories about the purpose of language: the point of discourse is to facilitate the coordination of our attitudes (1990).[10] It is similarly attractive to take the way in which our language can give voice to failures of coordination and refusals to coordinate as a basic part of linguistic meaning.

Beyond this, consider the explanatory commitments of expressivism. While many contemporary expressivists have quasirealist ambitions, quasirealism requires, minimally, that even if one accepts realist materials (objects, properties, truth) for the domain for which one gives an expressivist treatment, these materials play no role in explaining the relevant discourse (Blackburn 1993; and Dreier 2004). An expressivist may also wish to reject quasirealism (e.g., McPherson 2020), in which case her metaethical theory needs to be straightforwardly

[9]All this may seem to put us at odds with recent attempts to develop expressivism as a purely metasemantic doctrine (e.g., Ridge 2014; Chrisman 2016; and Köhler 2018). However, it is worth noting that some of dispute here is terminological. Ridge and Chrisman are explicit that by "semantics" they mean formal semantics. The formal account of meaning provided by such a semantics requires interpretation, however, and can be given a realistic or expressivist interpretation. Metasemantics for them includes such an interpretation, as well as metaphysical explanations of why, or in virtue of what, expressions have their meanings. (Köhler's view is a bit more complicated.)

We prefer the taxonomy found in (Baker 2021), on which "semantics" is used more broadly to include all aspects of the literal meaning of an expression. This includes the structural, compositional properties studied in formal semantics, but also the correct interpretation of the formalism. "Metasemantics" on such a taxonomy is reserved for explanations of *why* things mean what they mean. But for those who prefer other classifications can call the first proposal one in which disagreement is taken as primitive for the purposes of interpretative metasemantics, whereas it is explained in terms of functional roles of attitudes in our explanatory metasemantics.

[10]Also see Price (1990).

antirealist. Either way, the expressivist is committed to only allowing herself explanatory materials acceptable to an antirealist when it comes to the relevant domain.

Consequently, not all cases of semantic inconsistency are explicable in terms of attributing incompatible objects or states to the world—since some portion of discourse is not ultimately explained in terms of attributing objects or states to the world. This, as we have said, is the minimal commitment of expressivist analyses of normative thought and talk. We thus need a form of disagreement that doesn't reduce to incompatibility of referents.[11]

Taking disagreement as primitive also fits with another standard motivation for metaethical expressivism. Thought experiments like Hare's (1952/2003: 148ff.) Missionary and Cannibals and Horgan and Timmons' (1991) Moral Twin Earth seem to show two things. First, that our intuitions about disagreement in moral cases are different from our intuitions about disagreement with other kinds of terms (color-terms and natural-kind terms, respectively).[12] Our intuitions about disagreement in moral cases also do not seem to fit with standard referential semantics, which would typically predict of these cases that the presumptive disputes are merely verbal. One way of understanding the expressivist response to these cases is as an attempt to vindicate our intuitions by treating disagreement as primitive, as a distinctive part of meaning in addition to reference, rather than attempting to explain it in terms of reference.

Finally, taking disagreement as semantically primitive fits naturally with Greg Restall's (2005) proof-theoretic account of logic. Much of the paper will involve spelling out and adapting this account for expressivism. For now we can note that Restall's account depends on identifying rejection as a distinct kind of speech act, different from the assertion of a denial, and that it develops an account of logical consequence based on the incoherence of jointly accepting and rejecting certain packages of claims. It is natural to see this as one way of developing the Gibbardian proposal of taking disagreement as primitive.

[11] Even Schroeder's A-type expressivism does not reduce semantic inconsistency to expression of incompatible semantic objects, but expression of incompatible semantic objects plus *inconsistency transmission* of the attitudes (Schroeder 2008, ch. 3). But, at least at present, inconsistency transmission is explanatorily primitive, and so Schroeder's reductive move does not seem to lead to any ideological simplification beyond positing primitive disagreement relations.

[12] But see (Foot 1958/1959; Plunkett and Sundell 2014; and Dowell 2016) for criticisms of the thought experiments and the lessons commonly drawn from them.

3 Formalizing Disagreement

Let's say that Hare's missionary returns to the island of the cannibals.[13] Once again, they decide against eating him. (They've been worried his Christianity might be ingestible, ever since his explanation of the Eucharist.) For whatever reason, they instead want to discuss astronomical matters with him. They start to debate (a) and (b), whether the sun is a star. The cannibals, their cognitive abilities boosted by generations of brain-eating, naturally keep asserting (a). The missionary, in thrall to scriptural literalism, insists on (b). Treating 'not' as a force or attitude modifier, we can assign the following semantic values to their utterances:

(a′) Accept (the sun is a star)
(b′) Reject (the sun is a star)

More generally, if asserting 'P' expresses a commitment to Accept (p), then asserting 'not P' expresses a commitment to Reject (p). Intuitively, attitudes of accepting p and rejecting p disagree so, derivatively, the resulting commitments disagree. But in any case, since we are treating disagreement as primitive, we can stipulate as much. Two sentences disagree whenever one expresses Accept (p) and the other expresses Reject (p).

This is our basic stipulation. The question is whether we can use it to provide a constructive account of the logical properties of various sentences. Taking a small number of cases of disagreement as primitive, can we account for logical relations such as inconsistency and entailment in sentences of arbitrary complexity?

We will appeal to Restall's interpretation of the formal apparatus of Gentzen's sequent calculus to illustrate how one might go here. Note that this is a partially

[13]For the uninitiated, in the example of the missionary and the cannibals, R. M. Hare argues for expressivism on the basis of a way in which disagreement and translation involving normative terms seems unlike disagreement and translation involving descriptive terms. Let us say that by remarkable coincidence the words 'blue' and 'red' are found in the cannibals' language. But they typically apply 'blue' to what the missionary would call red, and 'red' to what he would call blue. It seems natural to say that the words mean different things in the two languages.

Let us say by further remarkable coincident, the cannibals use the word 'good', but they apply it to bold, cunning and ruthless warriors, whereas the missionary applies his word 'good' to those who are helpful, forgiving and self-effacing. The cannibals also tend to do what they call 'good'; they recommend 'good' acts; they admire the 'good' and despise those who are not 'good'. In this case, we are inclined to think that 'good' in the cannibals' language and 'good' in English have the same meaning, and the cannibals and missionary simply disagree about who is good. So while the lack of overlap in the extension to which the color terms are applied counts against translating them as meaning the same thing, the extreme difference in the extensions to which the evaluative terms are applied does not seem to count against their translatability. (See Hare 1952/2003 for more details.)

interpreted formal account; we're using it to limn the structure of these relations between the "semantic values" of various sentences, but it isn't by itself a full metasemantics.[14]

Gentzen's calculus provides a formal system in which arguments have not only multiple premises, but also multiple conclusions. A sequent line might look like this, for example:

$$p, q \vdash r, s$$

The sentences on the left hand side—the premises—should be read conjunctively. Those on the right should be read disjunctively. In other words, the above can be read as saying that an argument with premises p and q leaves open at least one of the conclusions r and s (though not necessarily both).

What to make of this? Restall offers the following interpretation. If we begin with an intuitive distinction between accepting a claim and rejecting one, we can understand the sequents as representing the incoherence of accepting the claims on the left while rejecting the claims on the right. In other words

$$p, q \vdash r, s$$

tells us that accepting both p and q rules out rejecting both r and s. If you accept p and q, you must, on pain of incoherence, leave open the possibility of accepting at least one of r or s. As Restall notes, incoherence can be explained in terms of truth, but need not be. Later, in **section 4** we will appeal to the idea of *discordance*, which we will identify with the *malfunctionality* of certain combinations of attitudes (as in Baker and Woods 2015). But there may be other alternatives as well. At this point what is important is that we can potentially explain logical consequence in terms acceptable to the antirealist.

A final point should be emphasized before moving on. On Restall's interpretation of Gentzen's calculus, '$\vdash$' means 'disagrees with' or "is incoherent with." But it also continues to play the normal role of the turnstile: it means 'proves'. So, '$p \vdash q$' may seem to simultaneously mean 'p disagrees with q' and 'p proves q'. But how can a claim disagree with what it proves?

The key here is that when we give '$\vdash$' the reading of "disagrees with," we must also read the sentences on the left-hand side as claims that have been accepted, and those on the right-hand side as claims that have been rejected. So '$p \vdash q$' does *not* say 'p disagrees with q'. What it says is 'Accepting p disagrees with rejecting q'. We can then say that on Restall's proposal, the more familiar reading–'p proves q'–follows from the novel one. Claim p proves

[14] See Yalcin (2014) for related discussion.

claim q because accepting p while rejecting q is incoherent. Claim q is proven because, so long as one accepts p, the only decided stance one can take towards q is also acceptance.[15]

With all this is place, we can verify the axiom of Gentzen's calculus:

$$A \vdash A$$

Remember, on Restall's interpretation of Gentzen's calculus, '$\vdash$ ' means 'is inconsistent with', or in our terminology, 'disagrees with'. The axiom can be read, then, as saying "Accepting A disagrees with rejecting A." Positing it as an axiom is simply to propose, as we have done above, that accepting and rejecting one the same claim is the basic case of disagreement.

How do we describe the logical properties of negation in this framework? This depends on what we take the logical properties of negation to be. A Pittsburgh-school relevantist will have a different interpretation than someone who accepts classical inference rules like *ex false quodlibet*. But if we accept something akin to classical logic, negation can and should be understood in terms of an illocutionary force modifier freely "swapping" acceptance to rejection and vice versa. That is, it changes an expression of a commitment to accept (we'll sometimes shorten this to 'commitment to accept') into an expression of a commitment to reject (analogously), and vice versa. The particle 'not', representing negation in the object language, can be seen as lexicalizing this operation of moving a claim from those we are committed to accepting to those we are committed to rejecting, and conversely.

This is captured through introduction and elimination rules for relevant connectives. In this context, sets of rules typically provide equivalences between patterns of accepting and rejectings. For example, here are the introduction rules for negation.

$$(\neg L)\left(\frac{\Gamma \vdash A, \Delta}{\Gamma, \neg A \vdash \Delta}\right) \qquad (\neg R)\left(\frac{\Gamma, A \vdash \Delta}{\Gamma \vdash \neg A, \Delta}\right)$$

The elimination rules do the inverse. We can read the left rule (indicated with L) as A is equivalent to accepting $\neg A$. And the right rule analogously.[16]

Capital letters in these rules represent claims. Greek letters represent sets of

[15] Additionally, it should be kept in mind that the right hand side of the Genzen calculus is read disjunctively when one interprets '$\vdash$' to mean 'proves'. '$p, q \vdash r, s$' means that either r or s follows as a conclusion from the premises p and q. On Restall's interpretation, this is because accepting p and accepting q while rejecting r and rejecting s is incoherent. So, as long as one accepts both p and q, one must either accept r or s.

[16] Typically, we suppress the side premises indicated with Γ and Δ in these readings to make the more easily comprehended.

claims. Note that "claims" should be understood broadly to include the meanings of any sentences that can stand in relations of disagreement with each other and various kinds of logical relations. Claims thus obviously include declarative sentences typically thought to express a proposition. But the claims in question could just as well be imperatives. Since disagreement and incoherence do not need to be cashed out in terms of incompatible truths, the introduction and elimination rules will hold so long as the claims in question are of the type in which relations of disagreement can be expressed using standard logical vocabulary. Given that "Close the door!" and "Don't close the door!" are both felicitous and intuitively express disagreement, the claims represented by our variables in the above rules could just as easily include imperatives.[17]

Returning to the introduction rules, we will introduce here the notion of equivalent commitments. Commitments are equivalent when they commit you to the same thing. To add some more precision, equivalence is transitive, symmetric, and reflexive, and if a set of commitments C is equivalent to another set of commitments C^*, then C disagrees with set of commitments D if and only if C^* disagrees with D. Note that two sets of commitments can be equivalent even if there is a commitment in one that isn't equivalent to any individual commitment in the other (exercise for the reader).

Given this explication, we have the following equivalences of commitments:

> REJECT (p) is equivalent to ACCEPT $(\neg p)$.
> ACCEPT (p) is equivalent to REJECT $(\neg p)$.

Given our base case of disagreement, these guarantee the inconsistency of p and $\neg p$.

This is only a partial account of the relation between acceptance, rejection, and the logical particles, however. What's wrong? First, we want rejection to be more than incompatible with acceptance—suspending judgment or having no opinion at all would suffice for simple incompability. (In Schlöder and Incurvati's terms, so far this could simply be *weak rejection*, and we want to specify that the rejection in question is strong.)[18] We also want rejection to positively rule alternatives out. Rejecting p should lead one to take the alternatives to accepting p;[19] or perhaps more accurately, it should lead one to take the alternatives to accepting p, or to revising one's picture of what those alternatives

[17]Thanks to a referee for asking us to be clearer on these points.

[18]See their (2017; 2019; and *forthcoming*). Note that we will not employ a primitive notion of weak rejection. The basic rejection relation for us is a strong one. This should not be understood as denying that weak rejection or weak assertion are genuine linguistic phenomena: Incurvati and Schlöder present independent evidence in favor of the existence of such speech acts. It is only that our semantic story here makes no use of these notions.

[19]This aspect of logical consequence has been emphasized in many places, most notably MacFarlane's (2004) (see also [Steinberger 2019]). It's been perhaps overemphasized, but nevertheless sets a *desiderata* for a successful account.

are (this is the familiar point that in reasoning the premises may force us to take the conclusion, but the unacceptability of the conclusion may also lead us to rethink the premises). In any case, we need some commitment with an alternative structure.

Second, and more importantly, the analyses of other logical connectives have "branching" structure. But how the branching works differs depending on whether the relevant connective is introduced into or eliminated from the left-hand or right-hand side of a given sequent–and whether a claim appears on the left- or right-hand side determines whether it is read as a case of acceptance or rejection. It follows that both acceptance and rejection of complex commitments are, in part, operationally defined by how they interact with branching commitments. Both of these points imply that we do not have a full account of the nature of rejection until we can spell out the role this commitment plays within a space of alternatives.

To illustrate, consider the conditional '$\rightarrow$'. The introduction rules are:

$$(\rightarrow L)(\frac{\Gamma \vdash A, \Delta \quad \Theta, B \vdash \Pi}{\Gamma, \Theta, A \rightarrow B \vdash \Delta, \Pi}) \qquad (\rightarrow R)(\frac{\Gamma, A \vdash \Delta, B}{\Gamma \vdash A \rightarrow B, \Delta})$$

Again, the elimination rules do the reverse (literally, read the intro rules upside down).

It's easy enough to capture $(\rightarrow R)$:

ACCEPT (p) and REJECT (q) is equivalent to REJECT $(p \rightarrow q)$.

What about $(\rightarrow L)$? Again, the introduction part is easy: it tells us that if two patterns of accepting and rejecting are incoherent, then so too accepting all the premises of those patterns with the conditional along with rejecting all the conclusions.

But the elimination part of $(\rightarrow L)$ is much harder. It might seem that we should define accepting a condition in terms of rejecting the consequent or accepting the conclusion. Notoriously, however, this kind of analysis will not do. It tells us that we cannot assert or believe a conditional claim without either rejecting the antecedent or accepting the conclusion. But we might accept a conditional without yet having decided to reject the antecedent or accept conclusion. We may be undecided at this point.

The elimination part of $(\rightarrow L)$ tells us that for all possible values for p and q, accepting a conditional of the form $(p \rightarrow q)$ disagrees with some further set of commitments C only if rejecting p disagrees with C and accepting q disagrees with C. So what is ACCEPT $(p \rightarrow q)$ equivalent to?

It cannot be equivalent to REJECT (p) *and* ACCEPT (q). Let C be {ACCEPT (p)}. In this case, ACCEPT $(p \rightarrow q)$ disagrees with C, because {REJECT (p), ACCEPT (q)} disagrees with {ACCEPT p}. But $p \rightarrow q$ is not inconsistent with p. So this predicts the wrong semantic relations. Nor can it be equivalent to REJECT (p) *or* ACCEPT (q), because as we have noted, we need not do either to accept $p \rightarrow q$.

We cannot explain the meaning of '$p \rightarrow q$' in terms of the commitments already posited. So we must propose an additional kind of commitment. Following Blackburn, we will call this commitment a *tree*. Trees are commitments to a pair of further commitments—commitments to not closing off all of a pair of further commitments. These further commitments can be acceptance, rejection, or further trees.

The formal properties of Tree are precisely those we need to capture ($\rightarrow L$): that accepting a conditional puts you in disagreement with a set of commitments if and only if accepting the consequent puts you in disagreement with the set and rejecting the antecedent also puts you in disagreement with the set. More precisely:

> Let C and D be sets of commitments and X and Y be commitments. $C \cup X$ disagrees with D and $C \cup Y$ disagrees with D if and only if $C \cup$ TREE(X, Y) disagrees with D.

With this is place, we can say what it is to accept a conditional:

> TREE (REJECT (p), ACCEPT (q)) is equivalent to ACCEPT $(p \rightarrow q)$.

Let us review briefly what we have done. Instead of treating the conditional as a truth function or set-theoretic operation, we are treating it as another illocutionary force modifier. It ties together two contents which would themselves express commitments into a commitment to leave one of two commitments open. Note that here we've moved from commitments to instantiate a pattern of, say, mental states to a commitment to undertaking commitments to mental states. Commitments, in other words, can stack.

In a way, this is to be expected, given our analysis of 'not'. 'Not', as we have said, modifies the force of assertions, changing them into denials, where the force of a denial is to rule out the commitment expressed by the relevant assertion. It would be strange if the other logical connectives did not similarly modify force rather than content. (It's hard to imagine how inferential roles could be vindicated.) So we are introducing another kind of illocutionary act, tying-oneself-to-a-tree. This is to commit to leaving at least one branch of the tree—the pair of commitments—open. In this particular case, it's a commitment to

not both accepting p and rejecting q. The remaining logical connectives can be similarly defined in terms of their relations to trees.

(To put aside a possible worry: yes, $(\rightarrow L)$ defines '$\rightarrow$ ' in terms of a conjunction. '$\rightarrow$' can be introduced just in case both '$\Gamma, A \vdash \Delta$' and '$(\Theta, B \vdash \Pi)$'. Why then is '$\rightarrow$' interpreted in terms of a commitment that has a disjunctive structure? Actually, the interpretation is a bit more complicated. Tying-to-a-tree, as defined so far, does not have a disjunctive structure. It is simply a commitment regarding a pair of commitments that is discordant with the set of commitments C just in case both members of the pair are. In other words, its disagreement relations have been defined conjunctively. It is true that insofar as one reasons from a tree to some conclusion, one will reason disjunctively, but that also reflects a feature of the logical formalism being deployed, albeit a different one from the interpretation of '$\rightarrow$ '. Remember, the right-hand, conclusion-side of a sequent is typically read disjunctively. Following Restall, we are capturing this by reading '$\vdash$' as 'disagrees with' or alternately 'is discordant with'. Accepting all of the left-hand formulae expresses a state of mind discordant with rejecting all of the right-hand formulae—or, to put it another way, one must change one's stance towards at least one right-hand formula on pain of discordance. Likewise, insofar as one reasons from a tree, one must leave open at least one branch, on pain of discordance. But the discordance relations are defined conjunctively.)

This is all nearly enough to verify *modus ponens*. All we need to add is a natural form of monotonicity for disagreement:

If C is a set of commitments that disagree with set D, and C is a subset of C^*, then C^* is set of commitments that disagree with D.

We want to show that '$p, p \rightarrow q \vdash q$'. On our Restall-inspired interpretation, this means that accepting (p) and accepting $(p \rightarrow q)$ disagrees with rejecting (q). ACCEPT $(p \rightarrow q)$ is equivalent to TREE (REJECT (p), ACCEPT (q)). Branch one is REJECT (p). REJECT (p) disagrees with ACCEPT (p), so by our intuitive principle, REJECT (p) disagrees with ACCEPT (p) and REJECT (q). ACCEPT (q) disagrees with REJECT (q), so ACCEPT (q) disagrees with ACCEPT (p) and REJECT (q). So accepting (p) and accepting $(p \rightarrow q)$ disagrees with rejecting (q).

3.1 A First Pass at Expressivism

"Alright," says the missionary, somewhat exhausted, "that's all very interesting, but what's that got to do with my problem? I'm trying to get these people to follow in the footsteps of Christ, and they believe, for example, that getting revenge is good!"

So we explain. Our account guarantees that he can argue with the cannibals.

On a referential theory of meaning, it is unclear that the missionary is entitled to translate them as saying that revenge is good. The term they use is applied to things like getting revenge, crushing enemies in battle, and lording it over others; whereas he applies 'good' to things like forgiveness, unconditional benevolence, and humility. There is so little overlap in the application, that it is hard to see what would justify assigning the same referent, and hence the same meaning, to each term.

To be clear, to say it is hard to see how the referentialist could justify assigning the same referent is not to say that they could not justify it. Only that it is difficult to see how such an account would work, and such an account looks like it would be complicated, at the very least.

The difficulty is that intuitively the cannibals and missionary are making claims that stand in contradiction. If they are contradicting each other, each means the same thing by 'good'. The referentialist has a harder time accounting for this. Why? She thinks that disagreement is (partly) explained by reference. So she needs some independent account of why the two uses of 'good' share a referent. Any such proposal will have to provide a theory of reference determination that is, at the very least, unusual. Normally the referent of a term is largely determined by how competent speakers are disposed to apply the term, or what causes them to apply the term, or something similar. So we need some explanation of why what normally determines reference fails to do so in this case, and instead something else does the job.

We have proposed treating disagreement as semantically primitive, however. And so we do not need to offer an *independent* account of why the two uses of 'good' co-refer, in order to explain why the missionary and cannibals disagree. An antirealist expressivist will simply deny that 'good' has a referent, and so no explanation of why they co-refer needs be given. A quasirealist expressivist will say the two uses of 'good' do co-refer, but reference for the quasirealist is a derivative semantic property. (Ultimately, reference for the quasirealist will be given a deflationary gloss, like truth, on which talk of reference is a device allowing for generalizations and semantic ascent.[20]) The quasirealist will thus say that the uses of 'good' co-refer *because* the cannibals and missionary disagree. In short, the expressivist has an easier time accounting for disagreement, because she does not need to give an independent account of why the uses of 'good' co-refer.[21]

[20] See (Field 1994; and Horwich 1998).

[21] A referee wonders if the referentialist could appeal to some sort of story like the expressivist's, only at a metasemantic level. 'Good' conventionally expresses some attitude, in virtue of which it has some reference, in virtue of which the sentences uttered by the missionary and cannibals disagree. Perhaps such an account could be developed. If it could be, with greater simplicity than an expressivist account, our motivation for expressivism would be defeated. The closest view in the neighborhood that we know of would be Ralph Wedgwood's (2007). We would only note that Wedgwood's view seems to depend on nonnatural properties to serve as the referent for normative concepts, and without this it would seem to be plagued by the kind of referential indeterminacy discussed in McPherson (2019). Eklund (2017) also provides

Our move of treating disagreement as primitive may not seem to accomplish much when we have disagreements which a referential account would predict anyway. The missionary rejects a proposition involving the property of *being a star* which the cannibals accept. Why all the work involving disagreement and trees when the truth-functional account would do just as well? But in the moral case, it is harder to say how the missionary and cannibals could be talking about the same property. Here's where the primitiveness of disagreement matters. We can simply posit that they disagree, even though there is no property at issue between them. They disagree about something non-propositional.

What is that? Here's a first pass. It will turn out to be inadequate, but it points in the right direction. They disagree about whether to get revenge. The cannibals accept getting revenge and the missionary rejects getting revenge. Here the value of treating disagreement as primitive and logical operators as force operators becomes apparent. Since our definitions of the logical operators and logical consequence make no reference to truth, they will work equally well when the content we accept or reject is non-propositional, when it is not truth-evaluable. We need only replace propositions (p, q, r) in the above definitions with act-types (ϕ, ψ, ξ), and we can say why 'revenge is good' is logically incompatible with 'revenge is not good' and can attribute a meaning to 'if revenge is good then holding a grudge is good' in a way that predicts the validity of *modus ponens* and *tollens*. The cannibals can say:

> If eating people is not good, then eating communion bread is not good. But you yourself acknowledge that eating communion bread is good. But then eating people is good. QED.

And we can explain why the argument is valid, without invoking truth or possible worlds or a common property under discussion.

"Wait," says the missionary, "are you claiming to have a solution to the Frege-Geach problem?" (Biblical literalism evidently does not inhibit understanding the philosophy of language the way it inhibits one's ability to do astronomy. Go figure.)

Our answer is that it depends what one means by the Frege-Geach problem. If it means providing a constructive account of how the logical connectives contribute to a sentence's meaning without appealing to truth values, almost. If it means something more, providing an account of all the relevant extralogical embedding relations—of what it means to hope that revenge is good, for example—then obviously not. But as far as answering Geach's initial objection goes, we are almost there.

---

an excellent exploration of the problems with reference facing normative realists, along with proposed solutions.

Of course, Blackburn and Gibbard were also almost there. And our first-pass stumbles in the same place they did. It tells us that 'Revenge is not good' expresses rejection of revenge. But that is wrong. The person who rejects revenge thinks revenge is bad. But holding that revenge is not good does not commit one to thinking that it's bad. One might think that it is neutral. Something for those who like it, but not worthy of encouragement or praise. The person who thinks that revenge is indifferent is not committed to ruling out revenge as an option—to rejecting it. This is the *problem of mere permissions* (Dreier 2006).

3.2 A Diagnosis and a Second Pass

It's worth getting clear on what the source of this new problem is. Here's our diagnosis: in the case of accepting and rejecting propositions, the stances of acceptance and rejection do not correspond to any part of the sentence we use to express the proposition itself. But, in the first pass account above, accepting or rejecting an act-type corresponds to the linguistic predicates 'is good' and 'is bad' and cognates. This mismatch is troubling.

Now, intuitively, both 'the sun is a star' and 'revenge is good' have subject-predicate structure. Negating either of them then suggests that the property picked out by the subject fails to apply to the subject. And we can treat this as rejecting that the property applies to the subject. This analysis works perfectly fine for the non-normative sentence. But in the normative sentence, we have already said that 'is good' corresponds to accepting, not a property like being a star. So how can we analyze negation for normative sentences? It should express rejection of the commitment to accept revenge. This, though, gives us a disunited account of negation. Sometimes 'not' expresses rejection of a content and sometimes it expresses rejection of accepting the content (i.e. the commitment to accept).

We think there is a relatively simple solution, however, drawing on Hare—and recent work inspired by Hare, from Nate Charlow (2014) and Andrew Alwood (2016).

Let ϕ be an act-type. Accept (ϕ) and Reject (ϕ) should correspond not to normative claims, but to imperatives—that is, accepting an act-type is intending it. Accept (*getting revenge*) and Reject (*getting revenge*) correspond to 'Get revenge!' and 'Don't get revenge!' on our proposal. There are no mere permissions when it comes to imperatives, at least not when we focus on their strictly literal meaning.[22] The negation of 'Get revenge!' doesn't simply present an alternative, but positively commands the addressee to rule vengeance out. To put the point another way, the negation of an imperative seems as decided as

[22]We recognize this is slightly controversial. See Portner (2016) for discussion and support of an alternative view.

the negation of nonnormative assertion. Imperatives thus fit better than normative assertions with acceptance and rejection of act-types. (Or perhaps it would be even better to say that the content of an imperative is a property: the commands "Be taller!" or "Be more massive than a star!" are not very sensible, but they are meaningful. The imperative would be satisfied, were the addressee to instantiate the property. So we should say that the content of the imperative is a property. Act-types can be thought of as a special subset of properties, the properties that tend to figure in the content of an imperative in the most standard cases.)[23]

To express mere permissions, we typically use assertions of deontic possibility.[24] 'You may get revenge', 'you are permitted to get revenge', and so on. These assertions conflict with the command 'Do not get revenge!', but in a manner that does not mandate performing the opposite act (Charlow 2014; and Alwood 2016).

Admittedly, we can in some cases use imperatives to express mere permissions. 'Close the door or leave it open.' 'Take a cookie if you want one.'[25] But two things should be noted. First, a disjunctive imperative, such as the first one, communicates permission to perform either disjunct. But it still issues an imperative: it commands (or advises, or requests) the addressee to perform one of the listed alternatives. If the addressee performs some action outside of the the alternatives presented in disjunction, they have failed to satisfy the imperative. While disjunctive imperatives do express a kind of permission, they still mandate some set of alternatives, which the simple assertion of deontic possibility does not.

Second, while 'Take a cookie if you want one' is naturally interpreted as expressing permission, we believe that this is because it pragmatically implies that one is permitted to take a cookie, rather than expressing such a permission in virtue of its literal meaning. In the same way, one could use a statement of fact to

[23]See Portner (2016).

Two further assumptions may be worth spelling out here, with the proviso that these assumptions are largely made for purposes of exposition, and to make the project of this paper more manageable.

First, we assume that deontic modals always require, recommend or permit *actions*. This may be unrealistic. Some sentences, such as, "There should be world peace," seem to require or recommend a state of affairs, without requiring or recommending that anyone bring the state of affairs about. If this is so, our theory only explains one part of the normative uses of modal terms, and the proposal must be extended to cover all of the normative uses of modal language. See Chrisman [2016] for more detailed discussion of this issue.

Our second assumption is that the content of an imperative is an act-type, rather than, say, an act-token. Our general thinking on this is simply that there are many possible token actions of getting revenge which would satisfy the relevant command. But we are not strongly committed to this, and if act-tokens were shown to be a superior content, we would be happy with that.

Thanks to a referee for asking us to be clearer on this issue.

[24]Cf. Thomasson (2020: 61).

[25]We should thank both referees for calling cases like this to our attention.

express permissions: 'You're an adult' or 'You don't work for this guy.' One could also use a question to do much the same, 'Why would anyone follow these orders?' 'Who cares if the Bible commands us to forgo revenge?'

We take it as most natural to interpret 'Take a cookie if you want one' as literally mandating a course of action, *conditional* on having certain desires. Strictly speaking, the imperative is flouted by the agent who wants a cookie but for some reason turns it down. It is simply that we naturally understand the speaker as leaving it up to the addressee. But on our view this is what we naturally understand the speaker as communicating, even if they are literally saying something else. We will simply say that while it is true that imperatives can be used to express permissions, our assumption is that statements of permission are unique in that (1) they express permissions as part of their literal meaning, and not simply through pragmatic implication; and (2) they express permissions without explicitly mandating some set of alternatives.

Returning to our main argument, we should also consider the connection between imperatives and claims of deontic necessity. 'You must ϕ' has been understood in the past just to be a kind of imperative (Kant 1785/1996; Hare 1952). If I tell you, 'You must get revenge', I seem to be telling you to get revenge in much the same way as if I were to issue the command. At the same time, the statement of deontic necessity is clearly more structured than a command. This was Hare's idea, 'You ought to ϕ' was a way of expressing a command to ϕ, but with extra semantic properties built in, namely that the command was addressed to every-one, everywhere who found themselves in similar circumstances, including the speaker (1952). We will develop this idea of a command with additional semantic properties, but in a different way, focusing on the fact that 'may', and 'must' are modals.

We have already committed to interpreting the logical connectives as force operators, as ways of lexicalizing certain kinds of commitments. While there is some question whether modal operators are part of logical vocabulary, strictly speaking (we will see in a moment that we probably have to deny that they are), they are at least close enough that there is some reason to think they deserve the same kind of interpretation: 'Might' and 'Must' should be understood as modifying commitments as well. In other words, there is some commitment, XXX, such that if 'Get revenge!' has the semantic value of

ACCEPT (*get revenge*)

then 'You may get revenge' expresses

XXX (ACCEPT (*get revenge*)).

This will plausibly do what we want. Remember, part of the problem, as we diagnosed it, is that negating a normative claim cannot be a matter of rejecting a content. There is no property whose predication can be denied, and rejecting the act-type itself is an overly decided state. So, negating a normative claim must be a matter of rejecting a commitment, not a content. By making the deontic modal express a commitment to another commitment, we are getting closer to a semantic value with the structure we want.

But will there be a problem in the descriptive case? Will we end up with descriptive sentences that are not decided enough? No. Remember that XXX has to modify the structure of the sentence expressing it. It forces us to add a modal operator. That means the corresponding commitment to a proposition

XXX (ACCEPT (*the sun is a star*))

should be the semantic value of the sentence 'The sun *might* be a star'. And that sentence is compatible with having suspended judgment on the matter. Of course, now we're committed to an expressivist treatment of epistemic modals as well. But there's already independent motivation for that position (Yalcin 2007 and 2011; Willer 2013; Incurvati and Schlöder 2019 and *forthcoming*). So instead of thinking of this as a further commitment, we could instead say that we are closer to a unified account of expressivism about deontic and epistemic modals.

But if we are already committed to expressivism about epistemic and deontic modals, then we are presumably committed to endorsing expressivism about metaphysical modals as well. Otherwise we would have to say that 'must' and 'can' have a radically different semantics in metaphysical contexts than epistemic and deontic ones. Fortunately, there is already non-referential story about metaphysical modals available. Borrowing from and slightly modifying Amie Thomasson's work (2007; 2013; 2016, and 2020), we propose the following analysis of metaphysical modals: they are devices for expressing semantic commitments within the object language.[27] This suggests a generalized form of modal expressivism: modal terms express metalinguistic commitments within the object language.[28]

[26] Matthew Chrisman (2016) also proposes treating deontic modals as taking imperatives as their prejacent.

[27] This involves modifying Thomasson's ideas, since her latter work explicitly adopts an inferentialist rather than expressivist understanding of metaphysical necessity claims. While Thomasson does in fact think that one of the uses of metaphysical modals is to *endorse* certain semantic rules (2020: 64ff.), they have this function in virtue of their connection to certain inference rules (2020: Chapter 4). Thanks to a referee for asking us to be clearer on this point.

[28] The idea that anti-descriptivists (whether inferentialists or expressivists) will need to provide a unified account of modal terms generally is not novel to us. Both Thomasson

This generalized account allows us to subsume existing proposals for expressivism about the normative and about 'might'-claims under a single picture. Alwood, as noted, has already proposed that 'may' functions to block or remove imperatives from the existing conversational context (2016). Incurvati and Schlöder propose treating 'might' as a way of blocking or removing assertions from the conversation context (2019). In short, if we see metaphysical modals as expressing *semantic* commitments (commitments about literal meaning) within the object language, we can see epistemic and deontic modals as expressing commitments regarding conversational appropriateness. Epistemic modals express commitment to certain constraints on assertability: to say 'Might p' is to express commitment to only taking on conversational commitments that are consistent with Accept (p) (that is, commitments which do not disagree with Accept (p)). To say 'May ϕ' is to express commitments regarding the prescribability of imperatives, that is to only taking on conversational commitments consistent with Accept (ϕ).

So we have an account of mere permissions: permission to get revenge does not rule out forgoing revenge; the permission rather rules out conversational commitments that would themselves rule out getting revenge. 'You may get revenge' functions to block, as it were, adding the command 'Don't get revenge!' to our conversational context. ('The sun might be a star' likewise blocks adding 'The sun is not a star' to the conversational context.)

That is the account we will offer in summary form. Before moving on to flesh it out, we need to finally discuss our metasemantics. What are these commitments? Where do conversational contexts come in? What is disagreement? This is especially pressing, because intuitively, 'It might be raining' disagrees with 'It is not raining'. So we need to explain why we are not committed to

$$Might\ p \vdash p$$

(2020: 14ff.) and Matthew Chrisman (2016: 194ff.) propose ways of unifying metaphysical modals with deontic modals, and acknowledge that there should be a unified account of modal language generally. Thomasson's proposal differs significantly from ours, however, in that she seems to regard deontic modality as the primary modal notion in terms of which other forms of modality should be explained (15). Ultimately the function of modal language is to convey rules, with metaphysical modals conveying semantic rules (16). Our proposal, as we will spell it out, is that no class of modals has any priority in understanding the others: all of them express different kinds of metalinguistic commitments.

Chrisman's proposal is closer to ours. He suggests that modal claims are "metaconceptual," meaning that they play the metalinguistic role of "affirming commitment... to the inferential relations between other words..."(194). He further suggests, as we will, that deontic modals play this kind of metalinguistic role with respect to imperatives. Chrisman's proposal is developed briefly, however, in a somewhat exploratory section of his book, and is presented in very technical language. Chrisman also only provides an account of necessity operators, not possibility operators. We take our account to be motivated by the same insights, but we aim to present a more fleshed-out account; and while technical language isn't fully avoidable, we hope to make the spell out the idea in more intuitive language.

since that would be bad.

4 Metasemantics of Disagreement

This section will offer a sketch of our views on the metasemantics of disagreement. This section will be guided by three goals. First, we must explain how utterances get their meaning, and what these meanings are. Second, we must say why it is that certain sentences disagree. Third, we must explain how it is that acceptance and rejection are the same thing whether the content being accepted or rejected is a proposition or a property (such as an act-type.

Our account of the meaning of assertions will be for the most part conventionally expressivist: assertions have the meanings they do in virtue of how they communicate our attitudes. In the basic cases, it is fairly simple to say what these attitudes are. To accept that p is to believe that p. To accept ϕ-ing is to intend to ϕ.[29] More complicated semantic values will be dealt with momentarily—we need a compositional recipe for identifying them, but that requires that we first identify what disagreement is. Equivalent semantic values are equivalent because the sentences express the same attitudes.

One complication needs to be addressed here, however. The semantic value of 'Get revenge!' is Accept (*getting revenge*), which we have identified with intending to get revenge. But 'Get revenge!' does not express an intention on the part of a speaker to get revenge. To explain this, we will take an idea from Gibbard: that the primary purpose of discourse is to help us coordinate our attitudes. Our view is that the semantic values of sentences should not be ultimately be understood in terms of expressed attitudes, but of those attitudes they invite interlocutors to coordinate around. 'Get revenge!' tells the addressee(s) to coordinate around an intention to get revenge. Assertions tell us which attitudes to coordinate around by expressing those attitudes, imperatives simply invite coordination without expressing that the speaker shares the attitudes as well.

Why the difference?Assertions invite coordination around attitudes for which there is more pressure to converge. Ordinary descriptive beliefs are about a shared world. Sincere communication on descriptive matters thus requires that I try to get interlocutors to coordinate on beliefs I share; otherwise I am trying to mislead them. As for *modal* assertions, the communicative function of language would be impeded if I kept inviting others to coordinate around metalinguistic

[29] A referee asks why accepting ϕ isn't instead an instance of preferring to ϕ. We have to admit we have no good answer. Readers who believe that preferences have enough rational structure to explain the disagreement relations of the relevant imperatives should feel free to replace our talk of intentions with talk of preferences. We intend for our proposal to be somewhat modular, so that various parts can be altered or replaced, without much change to the overall picture.

conventions I personally rejected. Descriptive and metalinguistic commitments are those about which we would expect people to push for coordination around the attitudes they themselves share. So it makes sense that there would be a speech act that invites coordination around an attitude by expressing the attitude.

In the case of intention, pressure is often against convergence. If Fred intends to marry Reza, things are actually better if no one else intends the same. So we have a speech act that invites intentions on the part of the specific addressee, without implication that the speaker will share them. You can advise Fred, "Marry Reza!" without any implication that you intend to do the same.

Given this picture of meaning, we can explain disagreement in terms of contravention of the expressed attitudes' functional roles (Baker and Woods 2015). A set of sentences disagree just in case the attitudes on which they invite conversational participants to coordinate cannot be jointly held without failures of those attitudes to fulfill their functional role.

With this in place, the meaning we have posited, made up of combinations of ACCEPT, REJECT, and TREE, can be thought of as a compositional story for specifying the attitude a given sentence invites coordination around. We started by defining the logical connectives in terms of their introduction and elimination rules–that is, in terms of their inferential role. Then, we identified the semantic values of the connectives in terms of inferential commitments, namely, ACCEPT, REJECT, and TREE. (This in contrast to, for example, a theory such as Wedgwood's [2007], which identifies the semantic value of a connective with the truth-function that would validate the connective's constitutive inferences.)

Now we have a recipe for saying which attitude a given sentence expresses (or invites). It is the attitude with the functional role which would lead the agent to satisfy the relevant inferential commitments, insofar as he or she is rationally coherent; it is the attitude whose functional role parallels the inferential role of the relevant sentence. ACCEPT (p) is equivalent to believing that p. Thus REJECT (p) is the attitude that (1) is functionally incompatible with believing that p, and (2) when combined with TREE(ACCEPT (p), C) commits the agent to C. TREE (ACCEPT (p), ACCEPT (q)) is equivalent to an attitude (1) the agent is committed to by believing p or by believing q, (2) which, when combined with reject (p) commits the agent to believing q, and when combined with reject (q) commits the agent to believing p, and (3) is functionally incompatible with rejecting both p and q. And so on for more complex trees. A similar story can be told for intention.

A few clarifications to make about this picture. First, this is not a higher-order attitude account of the meaning of complex sentences or of logical consequence (e.g., Blackburn 1984). A higher-order attitude is an attitude which takes another attitude as at least part of its content. That is not our proposal. ACCEPT,

REJECT, and TREE are not different attitude types, so that someone who asserts 'If the sun is star, get revenge!' is expressing an attitude of deciding-between towards an attitude of rejecting of the sun's starhood or an attitude of accepting revenge. Rather TREE (REJECT (*the sun is a star*)), ACCEPT (*get revenge*)) should be understood as a compositional way of specifying functional role. The inferential role of the structured commitment TREE (REJECT (X), ACCEPT (Y)) is a compositional way of telling us how the speaker's psychology should be with respect to X and Y, and how it should utilize those contents when other attitudes are also present.

But what are the attitudes in question? Which attitude has the functional role specified by REJECT (*the sun is a star*)? The English-language name for the attitude is believing that the sun is not a star. TREE (REJECT (*the sun is a star*), ACCEPT (*revenge*)) is an intention to get revenge conditional on the sun being a star. English is a logically structured language, and so there is likely no name in English for these attitudes that does not identify them on the basis of a logically complex content. This was already implicit in our story of the logical connectives: as we said, REJECT (p) is equivalent to ACCEPT ($\neg p$). But if ACCEPT (p) is equivalent to believing that p, then REJECT (p) must be equivalent to believing that $\neg p$.

Are we saying anything beyond the banality that 'the sun is not a star' expresses the belief that the sun is not a star? We do not wish to be global expressivists, and so it is our ardent hope that when it comes to ordinary descriptive claims our view is composed almost entirely of banalities. As we made clear earlier, our aim was never to provide a reductive story about the logical connectives—no account does that. What we have been trying to do is show how an expressivist-acceptable account of these connectives is possible. We did this first by identifying commitments that stood in relations of incompatibility, and then explained these commitments and their incompatibility in terms of the functional roles of beliefs and intentions. Explaining the meaning of REJECT (p) in terms of the belief that $\neg p$ only undermines that project if we have to appeal to the content, $\neg p$, in order to explain the functional role. The account would then indeed be circular. But we maintain the reverse. The functional properties of the state explain why it counts as a rejection of p, rather than an acceptance, and this explains why it is a belief that $\neg p$. (This is a position that will be familiar from conceptual role semantics about logical terms; see for example, [Wedgwood 2007].) The upshot is that we will explain what 'May ϕ' means, for example, and how it can stand in logical relations to other sentences, without presupposing that 'May' contributes normative content–or any content at all–to the sentence in which it occurs.

An important problem remains. ACCEPT and REJECT are supposed to specify functional roles of attitudes. But already we've had ACCEPT specify two distinct attitudes—belief and intention—with clearly different functional roles. Moreover, we promised to explain the meaning of our logical vocabulary in terms of

how they modified basic attitudes. So REJECT (p) and REJECT (ϕ) presumably specify some modification of belief and intention, respectively. But on what basis can we assume this is the same modification? And if it isn't the same modification, in what sense in the specification of attitudes compositional?

The answer is that our stance operators—ACCEPT, REJECT, TREE—are not meant to specify all of the functional properties of the relevant attitude. In fact, we do not want them to specify all functional properties. These stance operators are meant to account for the logical relations that hold between various sentences expressing the relevant attitudes. But the vast majority of the reasoning we engage in is not deductive. If we built every functional property into the meaning of our sentences, we would end up with the unattractive position that all reasoning is deductive and all truths are analytic.

With this in mind, we can say how 'ACCEPT', 'REJECT', and 'TREE' have univocal meanings, whether the attitudes denoted by them take acts or propositions as their object. The idea is that both intending and believing have enough rational structure in common that we can identify a particular state of intending or believing by specifying it as the attitude which plays a certain deductive, or deductive-like, in the case of intentions, role. The reasoning in question, specifically, is that of identifying alternatives, and either selecting an alternative or ruling out alternatives, a kind of reasoning we engage in both when forming beliefs and when deciding what to do (though further aspects of that reasoning, such as why two stances are considered alternatives, or the basis on which alternatives are selected or ruled out, might be very different). Moreover, there is rational interaction between our beliefs and intentions. This is important because it allows us to appeal to a generic notion of a tree. Trees can take stances toward propositions as their objects, or stances toward acts, or a mixture.

In short, ACCEPT, REJECT, and TREE should be thought of as terms in a compositional system for specifying *some* of the functional properties of the mental state expressed by or invited by some sentence. The functional properties in question are those which are deductive-like: those which can be specified as more or less complex combinations of laying out alternatives, selecting from them, or ruling some out. Importantly, the combinations of deductive-like functional roles for an attitudes can be specified by the logical form of the sentence expressing the attitude. These deductive-like properties are the functional properties that beliefs and intentions have in common, even though there are a number of other functional properties they obviously do not have in common. This suffices for sameness of meaning of our technical vocabulary.

One might worry that we have assumed that specifying the broadly 'deductive' role of an attitude is sufficient to fully individuate the attitude type—that, for example, there are not two different attitudes which might both play the same role in logical reasoning that Reject (p) plays on our account. Actually we need no such assumption. In fact, even if all cases of rejecting that p are cases of

believing that $\neg p$, this would not ensure that 'Reject (p)' denoted a single type of mental state. Why not? There is no guarantee that belief itself is a unified mental type.

Consider the following proposal, made in a different context: "We do not hold that belief is a unitary phenomenon. Rather than a simple sort of cognitive attitude, 'belief' may be a convenient label that lumps together a host of slightly different phenomena" (Boudry and Coyne 2016: 602).[30]

We have no opinion on whether this is the case, but if it is, 'belief', to be a convenient label, must still communicate useful information about the subject's mental state, even if it does not fully specify the attitude in question. In a similar way, knowing that a speaker is in a state with the broadly deductive role of REJECT (p) would provide important information about their state of mind, even if there are details about their exact stance toward p that are left out. To fill in these details we may need to learn what they take to be evidence for and against p, to observe their behavior, to learn about how they came to reject p, and so on. But again, not all aspects of the role an attitude plays in our reasoning will be captured by the semantic properties of the sentence expressing that attitude.

A final objection. Why can't we say, in English, 'If go to the store, then get revenge!'? This sentence would presumably express TREE (REJECT (*go to the store*), ACCEPT (*get revenge*)). And yet the sentence is gibberish. So there must be restrictions on the ways in which our force modifiers can apply to act-types (and other properties). But we have given no account of why there would be such restrictions. Furthermore, that the combinations of TREE, ACCEPT, and REJECT that are applicable to propositions cannot be applied to act-types suggests that these force modifiers are not the same thing when applied to the latter as when applied to the former (since they have different compositional properties).

The short answer is we do not know why the relevant sentence is gibberish, but this is not actually an objection to our view, as our view at present makes no predictions about the conditional, 'If go to the store, then get revenge', at least not if this is interpreted as a natural-language conditional. We have aimed to give an account of the material conditional and other Boolean connectives. At present we do not know how to handle natural language conditionals. The suggested combinations ACCEPT, REJECT, and TREE are meant to provide the semantic values of *Boolean* connectives (or something close enough to Boolean, anyway).[31]

[30] For a similar position on belief, see (Schwitzgebel 2002 and 2013); Michael Ridge also argues, in the context of his hybrid-expressivism, that normative beliefs are multiply-realizable (2014: 118ff., and 194ff.).

[31] A complication: depending on which rules we accept, the semantics here might be somewhat different from standard Boolean connectives. Perhaps the conditional is slightly weaker than the classical conditional. The point is simply that we can express the idealized con-

An imperative cannot be the antecedent of an 'if... then' conditional in English, that is clear. An imperative can, however, be the antecedent of '→', which is what our account actually predicts. The source of the objection is thus that we have been treating the material conditional as equivalent to the English-language 'if...then'. This obviously involves considerable idealization. Stepping away from this idealization, TREE (REJECT (*go to the store*), ACCEPT (*get revenge*)) is expressible in English, in the form of the disjunctive command, 'Stay away from the store or get revenge!'[32] These are logically equivalent to 'go to the store → get revenge'. Consequently this sentence is also expressible, though it would only be expressed in somewhat artificial conditions. Admittedly this means we have only shown how expressivists can validate a somewhat idealized proxy for *modus ponens*: but this is true of all existing expressivist solutions to the Frege-Geach problem, so far as we know. We are starting by trying to show that expressivists can vindicate basic logic. More complex expressions and their logical relations will have to come next.

Furthermore, there is reason for optimism about natural-language conditionals within our framework. Conditionals are standardly treated as modal claims or closely related to modal claims (e.g., Kratzer 2012). It is natural in the proposed framework to investigate, then, whether natural language conditionals can also be understood as expressing metalinguistic commitments.

To summarize, we have explained why a set of conversational commitments is in disagreement. Each commitment is a commitment to a certain attitude. Commitments disagree just in case the functional roles of the attitudes are incompatible. The logical form of a sentence thus specifies very general functional properties of the attitude, in a way that parallels the sentence's inferential role. These functional properties are common to both theoretical and practical reasoning: namely, comprehending alternatives, ruling out some of those alternatives, and accepting from among those alternatives that remain. This is only an account of Boolean connectives, which means our account validates a highly idealized version of *modus ponens*. But we don't mean to build Rome in a day.

5 Metalinguistic Commitment and Reasoning

With our story about the nature of disagreement in place, we are in postion to explain why $Might\ p \nvdash p$. It starts by denying a standard position within the literature. The position goes like this. 'It might be raining' does not express a belief in any particular proposition (there are no might-propositions), but rather

ditional in terms of inferential commitments—however that is to be done—without denying that the ordinary English-language conditional is to be handled differently.

[32]'Don't go to the store or get revenge' is most naturally read as giving the negation widest scope. We use 'Stay away from the store or get revenge!' as an equivalent command without any confusing scope ambiguities.

expresses a belief-set that is consistent with the claim that it is raining. (Yalcin 2011; and Willer 2013).

The position is subject to obvious counterexamples. It is admittedly infelicitous to say, 'It is raining but it might not be raining'. However, it is perfectly fine to assert, 'I believe it's raining, but it might not be' or 'It might be raining, but I don't think it is'. These sentences do not sound self-contradictory, Moore-paradoxical, or otherwise self-undermining. One might object that clauses such as 'I believe it is raining' do not express a belief that it is raining, but rather a belief that I believe it is raining. This is correct, but it still must be the case that in order to sincerely express these sentences I must be self-deceived about my own beliefs, which should lead the claims to sound self-undermining. And in any case, it is extremely counterintuitive to think that there is some sort of functional incoherence in believing something, while also accepting that one might be mistaken. Each of us believes many things we think we might be wrong about.

The position also has unhappy consequences if we try to generalize it to the deontic case. If 'might' expresses consistency with a belief-set, 'may' should express consistency with a set of intentions (or preferences, or other practical attitude). But consider the sentence 'One may refrain from closing the door'. It should express–or better yet, *invite*–a set of intentions consistent with the intention not to close the door. In other words, accepting the permission is incompatible with still intending to close the door. But this would make permissions into obligations to either perform the permitted act or remain undecided. But that's obviously not what permission is. Note as well that it is fine to say, 'You may leave the door open, but if you intend to close it that's fine too', as is, 'Working from home is permitted, but I intend to work at my office today'.

Epistemic and deontic possibilities are not primarily concerned with what is consistent with the speaker's, or anyone else's personal attitudes. Rather, they govern attitudes shared by the group: specifically 'might' and 'may' deny that listeners should converge on rejecting the prejacent.

If I assert, 'It is raining', I can normally assume that my conversational partners believe that it is raining, unless they object. After all, the statement is supposed to present the expressed attitude as one to converge on. How do they reject the claim? The obvious way is to say 'It is not raining'. But notice that this now presents the belief that it is not raining as a state to converge on. What if your objection is weaker? You are, for example, undecided about whether it is raining. In that case one can say, 'I don't believe it's raining', or 'The evidence about that is inconclusive'. This signals that we should not make the claim part of our conversational common ground, without making its negation part of our common ground instead. One can also say, 'It might not be raining'. The epistemic modal functions to block an assertion without asserting the contrary

(Incurvati and Schlöder 2019).[33]

The deontic modal seems to have an exactly parallel role, but with imperatives. There are numerous ways of objecting to the command 'Close the door!' One can issue the addressee an opposite command, 'No, leave the door open!' One can make various implicit challenges, 'I'm not going to', or 'You're not my boss'. But one can also say: 'Bob may refrain from closing the door' (Charlow 2014; and Alwood 2016). In other words, assertions of deontic possibility block imperatives, without issuing a counter-imperative.

This fits with what was noted previously. The person who utters, 'It might be raining, but I'm pretty sure it's not', is obviously not objecting to the belief that it is raining. He is instead objecting (perhaps preemptively) to an insistence that we converge on this belief. (Most likely, he does not regard the grounds of his belief as sufficient to insist that everyone shares it.) Similarly, the person who says, 'You may leave the door open', is not objecting to the intention to close the door, but the insistence on the intention.

This is explained in our framework if we interpret modals as expressing metalinguistic commitments. Asserting that p adds ACCEPT (p) to the conversational common ground, which means that other speakers must either dissent or converge on accepting p, on pain of failure to be cooperative communicators. If 'might $\neg p$' expresses commitment to norms of assertion on which claims inconsistent with $\neg p$ are unassertible in the current context, this will block the implicit insistence on convergence. *Mutatis mutandis* if 'may $\neg\phi$' expresses norms regarding the prescribability of imperatives. We should note that the conversational norms in questions should be understood as norms governing communication so long as strategic concerns do not become paramount. We may, for example, think a certain claim is unassertible due to politeness or deference, or because it would reveal information we would rather not reveal, or because everyone is becoming bored of the argument, but these are a cases where the normal goal of coordinating attitudes–which can also be thought of as collaborative reasoning–become secondary to other concerns. It is likely, moreover, that these forms of strategic communication are parasitic on the more standard, collaborative case.

As it is, however, the account is incomplete. As expressivists, we cannot accept unexplained appeals to metalinguistic *norms* or to normative notions such as 'assert*ibility*'. We will follow Gibbard again: to accept metalinguistic norms is ultimately to have stable metalinguistic intentions–intentions about how to speak oneself and what forms of speech to object to from others.

What is the content of these intentions? Here is a proposal. 'Must A' expresses an intention not to object to assertions or prescriptions expressing commitments equivalent to those expressed by the prejacent. 'Might A', as the dual of 'Must A' should then consist of an intention to object to assertions or prescriptions

[33]This actually simplifies their view a bit.

expressing commitments equivalent to those expressed by the *rejection* of the prejacent. In case of both 'Must' and 'May' these intentions should be understood as intentions to speak and publicly reason this way conditional on the intention of communicating with the default Gibbardian aim, rather than some strategic aim.

With all this in place, do we get the inferential relations right? We need the following to be true:

$$Must\ p \vdash p; \quad p \vdash Might\ p; \quad p \nvdash Must\ p; \quad Might\ p \nvdash p$$

Accepting *Must* p amounts to intending not to object to anyone asserting that p, and intending to object to any argument that could not add acceptance of p as a premise without incoherence. Is there functional incoherence when this is combined with the belief that $\neg p$? It is. First, consider an important feature of the nature of the intentions. As noted above, this communicative intention is not strategic. The first intention intention is fully characterized as an intention not to object to p when conversing for the purpose of coordinating attitudes. But if one believes $\neg p$, this aim is guaranteed to be frustrated. You cannot reach of the goal of coordination by refusing to object to p, so long as you do not believe p. 'I must ϕ' expresses, likewise, an intention not to object to commands or advice to ϕ, when communicating cooperatively. But such an intention must be unsatisfied so long as I lack the intention to ϕ. In short, the functional role of these metalinguistic intentions is guaranteed to be frustrated so long as their possessor does not have the commitments expressed by the prejacent. Again, the intentions in question should be understood as conditional: we intend to communicate in this way when we have the intention of communicating non-strategically.

At the same time, as we have noted above, one can coherently believe p or intend to ϕ while intending to object to adding the relevant commitment to the conversational common ground. For whatever reason, one has one's own views, while acknowledging that intersubjectively things remain inconclusive.

Tidying things up, we can introduce a new conversational commitment to be expressed by our modal operators. This name will serve really as a way of summing up the pair of metalinguistic intentions described above. On our account, we can think of 'must A' as expressing a commitment to metalinguistic norms on which acceptance of A should be assumed to be common ground in the conversation. This then gets cashed out as an intention to engage with interlocutors in a way that amounts to treating acceptance of A as already implicitly part of the common ground. So, we will call the relevant intention expressed by 'must A':

ACCEPT (COMMON (ACCEPT (A)))

With this in place we can say:

> ACCEPT (*Must X*) is equivalent to ACCEPT (COMMON (ACCEPT (*X*))).
> REJECT (*Must X*) is equivalent to REJECT (COMMON (ACCEPT (*X*))).
> ACCEPT (*Might/May X*) is equivalent to REJECT (COMMON (REJECT *X*))).
> REJECT (*Might/May X*) is equivalent to ACCEPT (COMMON (REJECT (*X*))).

> For any two commitments X and Y, X is equivalent to Y if and only if Accept (Common (X)) is equivalent to Accept (Common (Y)) if and only if Reject (Common (X)) is equivalent to Reject (Common (Y)).

> Let X be a conversational commitment and C be a set of commitments. In this case, if X disagrees with C, then ACCEPT (COMMON (X)) disagrees with C.

From these rules the following inferences hold:

$$Must\ p \vdash p;\ \ p \vdash Might\ p$$

The unwarranted inferences, on the other hand, do not hold.

5.1 Negation, Mere Permission, and Semantic versus Pragmatic Disagreement

It is worth noting that in order to solve the negation problem, we had to propose that modal (and hence deontic) statements had a significantly different role than sentences without normative vocabulary. If the latter express attitudes, the former have the job of expressing principles regarding the expressibility of attitudes. Or better yet, the latter have the job of coordinating attitudes, the former the job of coordinating our practices of coordination. (This means that the attitudes expressed by modal claims are genuinely higher-order. We're okay with this.)[34]

[34]A referee notes that Yalcin (2011: 307-9) objects to explaining what it is to think that it might be raining in terms of higher-order attitudes. However, as Yalcin states the objection, he is clearly objecting to accounts which treat the thought that it might be raining as a higher-order *belief*. On our account, what is higher-order is the metalinguistic intention. Nonetheless,

How this plays out, in concrete terms, is that while the connectives express commitments of individual reasoners, possibility operators do not commit individual reasoners to anything; rather, they express interpersonal commitments about where and where not to insist on convergence with others. It is true that all assertoric speech invites convergence, on our view. But in nonmodal assertions, it does so by inviting each individual participant in the conversation to take up the same commitment. In the case of 'might' or 'may', however, it leaves open that individual reasoners come to whatever conclusions they want; it simply commits to leaving the commitment expressed by the prejacent open to conversational participants.

This distinction is necessary if we want to allow for mere permissions, while accounting for logical consequence in terms of rejection. On Restall's interpretation of '$\vdash$' marks the functional incoherence of a package of acceptances and rejections. But '$\vdash$' also has its standard meaning of 'proves'. A follows from B because Accepting B while Rejecting A is incoherent. But this leaves us with a difficulty. 'May ϕ' needs to rule out $\neg\phi$ in some way, and 'might p' needs to rule out $\neg p$ in some way; but not in a way such that $\Diamond x \vdash x$, otherwise we have not solved the problem of mere permissions. So we introduce the following distinction. Logical connectives express psychological commitments—commitments in how to reason individually. Modal operators express metaconversational commitments—commitments about how to reason collaboratively. '$\vdash$', denoting functional incoherence, shows us what conclusions individual reasoners must accept. As a metaconversational commitment, 'may/might' gives minimal instructions on what individual reasoners are to do.

Standard realist accounts solve the problem of mere permissions by appealing to quantificational structure. On these, 'might/may x' tells us that there is at least one possibility, within the relevant restricted domain, in which x. This is compatible with $\neg x$ in at least one other possibility, and so both are left open. We can see our solution to the problem of mere permissions as implicitly appealing to the same kind of quantificational structure. Gibbard's hyperdecided agents (1992; and 2003) can obviously be seen as a device for assimilating resources from possible-worlds semantics into a psychologized, expressivist-friendly theory of meaning. Modals could be understood, on this picture, as telling us which hyperdecided agents could be sincere participants in our conversation. In other words, the semantic value of 'might/may x' could be thought of as expressing that at least one hyper-decided agent in the set of agents expressed by 'x' has beliefs and plans consistent with those expressed by the conversational common ground. 'Must x', alternately, tells us that only hyperdecided agents in the set expressed by 'x' have beliefs and plans consistent with the common ground.

one of his objections would clearly apply to our account: on our account, a dog cannot think that you might be about to give him a bone (*ibid.*) Again, we are okay with this. We think many of the mental states attributed to animals in everyday conversation involve considerable anthropomorphism, or else are not intended literally. Yalcin's other criticisms of higher-order attitudes do not seem to apply to our account.

This raises a question, however, whether our account treats the conflict between 'may/might x' and '$\neg x$' as pragmatic or semantic. The answer is complicated: for an expressivist, the distinction between pragmatics and semantics is less clear cut.[35] Our view is that the conflict in question is roughly equivalent to a conflict of *conventional* implicature. On orthodox semantic theories, a conventional implicature is a claim implied by the conventional meanings of the words; but it is not part of the truth conditions or entailed by the truth of what is uttered. 'But' makes the same contribution to the truth conditions of a sentence as 'and'; but unlike 'and' it further communicates some sort of contrast between the two conjuncts. 'Mary is poor but honest' has the same truth conditions as 'Mary is poor and honest', but the former sentence, unlike the latter, implies some sort of tension between poverty and honesty. Importantly, implicatures such as these are normally treated as semantic, as they are part of the conventional meaning of the words, rather than what one infers the speaker to be communicating on the basis of context.[36] At the same time, they are unlike paradigmatically semantic properties, in that they play no role in determining an utterance's truth conditions.

Our account cannot, for reasons that should be clear by now, treat truth-functionality as the primary or paradigmatic feature of semantic properties. Paradigmatically semantic properties, on our view, are those that determine relations of disagreement or discordance. These are properties that determine a kind of disagreement which explains relations of entailment; this kind of disagreement explains relations of entailment, moreover, by ruling out certain inferences. The disagreement between 'may/might x' and '$\neg x$' is not this kind of disagreement. 'May/might x' does not entail x. Our account explains this by positing that the modal claim expresses a commitment to linguistic conventions regarding the unassertibility of '$\neg x$', which is still compatible with individually inferring that $\neg x$. So this is not a *paradigmatically* semantic form of disagreement. At the same time, the disagreement in question is based on the conventional meaning of the words 'may' or 'might'. This was the contrast introduced earlier between 'It might not be raining' and saying 'I don't believe it is raining' in response to someone's claim that it is raining. The latter expresses the same kind of disagreement–disagreement with adding *it is raining* to the conversational common ground–but does so as a matter of pragmatics. So the conflict between 'may/might x' and '$\neg x$' is semantic, in the sense that it is a conflict based on conventional meaning, rather than contextually supplied implications.

[35] We would like to thank one of our referees for calling this point to our attention, which helped us to develop a more principled account of the kind of disagreement involved in modal expressions.

[36] See Davis 2019: §2.

6 The Final View

We have now finished setting out our view. The novelty of it is that we treat both deontic and epistemic modals in a unified way. Both express commitments about our commitments—such as commitments to not insist on convergence on a particular belief or intention. These metaconversational commitments allow us to solve the problem of mere permissions.

This is admittedly a first step. There are several obvious lacunae. The semantic story we have told about modals remains minimal and needs fleshing out. We need moreover to show more systematically how an account of alethic modals such as Thomasson's will fit in with our proposal, in a way that goes beyond simply noting the natural affinity. Our account of logical vocabulary needs to be expanded to make sense of quantifiers and of real conditionals, rather than the simplified Boolean connective.

What do the cannibals mean by 'good'? A simple answer, following the inspiration we have taken from Gibbard, is to analyze evaluative phrases in terms of deontic ones: treat 'good' as synonymous with 'desirable' and 'desirable' as meaning 'may be desired'. For now that will be our official story.

We think, though, there is reason to look at a more complicated one. Stephen Finlay argues persuasively that the meaning of 'good' should be understood in terms of conditional probabilities (2014). The actual semantics are too complicated to present here, but the rough idea is that 'X is good' communicates that X makes more likely the achievement of some contextually relevant goal. One advantage of this account is that it allows a unified semantics for both "predicative" and "attributive" uses of 'good'.

The expressivism presented here has been motivated by a sense that if the deontic 'must' and 'may' are given an expressivist analysis–as they must be for a metaethical expressivist–we are also committed to expressivism about other modal expressions. Expressivism about the deontic turns out to be a special case of expressivism about modals generally. It seems natural then to consider that expressivism about 'good' is really a special case of expressivism about conditional probabilities more generally.

We call this account expressivist, though the acute reader might notice that it has ceased to resemble the classic expressivism of Blackburn and Gibbard. Rather, it seems to sit between their expressivism, which focuses on states of mind, and more inferentialist approaches. As noted above, it differs from the former in that the fundamental bit of machinery, commitments, are social, not mental, even though mental states do play a role in the view. And it differs in the latter because the commitments we make use of do not instruct us to infer or justify anything; rather, they're just commitments to be. If you think the resulting view isn't expressivist enough, so be it; we disagree. But names aren't

really the issue anyways.

We note that our view is also compatible with a variety of motivations for expressivism and a variety of further developments. We motivate our position here through classic apparent anomalies of normative disagreement—Hare's Missionary and Cannibals case, for example. But one could just as easily be brought to this position by a rejection of moral realism. The proposal can be developed into a form of quasirealism, or one may insist on a more revisionary antirealism. The project here is to show how Gibbard's idea of treating disagreement as semantically primitive can be developed into a constructive account of logical and modal vocabulary, and how this can solve the problem of mere permissions.[37]

References

Alwood, Andrew (2016). "Non-descriptive Negation for Normative Sentences." *Philosophical Quarterly* 66 (262): 1-25.

Baker, Derek (2021). "Quasirealism as Semantic Dispensability." *Philosophical Studies* 178 (7): 2313-33.

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

Blackburn, Simon (1984). *Spreading the Word: Groundings in Philosophy of Language.* Clarendon Press.

Blackburn, Simon (1988). "1Attitudes and contents." *Ethics* 98 (3):501-517.

Blackburn, Simon (1993). *Essays in Quasi-Realism.* Oxford University Press.

Brandom, Robert B. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment.* Harvard University Press.

Charlow, Nate (2014). "The Problem with the Frege-Geach Problem." *Philosophical Studies* 167/3: 635-65.

Chrisman, Matthew (2008). "Expressivism, Inferentialism, and Saving the Debate." *Philosophy and Phenomenological Research* 77: 334-58.

Chrisman, Matthew (2016). *The Meaning of 'Ought': Beyond Descriptivism and Expressivism in Metaethics.* Oxford University Press.

[37]Acknowledgments: Research appearing in this paper was substantially funded by a General Research Fund grant from the Research Grants Council of Hong Kong SAR, China (Grant Number LU13601218). We would like to thank the audience at the 2016 Lingnan Expressivism Conference at Lingnan University, Hong Kong, for helpful comments, suggestions, questions and criticism. We would also like to thank the two anonymous referees of this paper for the same.

Davis, Wayne (2019). "Implicature." *The Stanford Encyclopedia of Philosophy.* Fall 2019 edition. E.N. Zalta (ed.). URL = <https://plato.stanford.edu/archives/fall2019/entries/implicatu

Dowell, Janice (2016). "The Metaethical Insignificance of Moral Twin Earth." Oxford Studies in Metaethics 11: 1-27.

Dreier, James (2004). "Meta-ethics and the Problem of Creeping Minimalism." *Philosophical Perspectives* 18: 23-44.

Dreier, James (2006). "Negation for Expressivists: A Collection of Problems with a Suggestion for Their Solution." *Oxford Studies in Metaethics* 1: 217-33. R. Shafer-Landau (ed.). Oxford University Press.

Dummett, Michael (1974). "The Justification of Deduction." In Michael Dummett (ed.), *Truth and Other Enigmas.* Oxford University Press.

Eklund, Matti (2017). *Choosing Normative Concepts.* Oxford University Press.

Field, Hartry (1994) "Deflationist Views of Meaning and Content." *Mind* 103 (411): 249-85.

Foot, Philippa (1958/1959). "Moral Beliefs." *Proceedings of the Aristotelian Society: New Series* 59: 83-104.

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

Gibbard, Allan (2003). *Thinking How to Live.* Harvard University Press.

Hare, R.M. (1952/2003). *The Language of Morals.* Clarendon Paperbacks.

Harman, Gilbert (1986). *Change in View.* MIT Press.

Horgan, Terence & Mark Timmons (1991). "New Wave Moral Realism Meets Moral Twin Earth." *Journal of Philosophical Research* 16: 447-65.

Horwich, Paul (1998). *Truth.* Oxford University Press.

Incurvati, Luca, & Julian Schlöder (2017). "Weak Rejection." *Australasian Journal of Philosophy* 95 (4): 741-60.

Incurvati, Luca & Julian Schlöder (2019). "Weak Assertion." *Philosophical Quarterly* 69 (277): 741-70.

Incurvati, Luca & Julian Schlöder (*forthcoming*) "Inferential Expressivism and the Negation Problem." *Oxford Studies in Metaethics* 16. R. Shafer-Landau (ed.). Oxford University Press.

Kant, Immanuel (1785/1996). *The Groundwork of the Metaphysics of Morals.* In M.J. Gregory (ed. and trans.). *Kant's Practical Philosophy.* The Cambridge Edition of the Works of Immanuel Kant. Cambridge University Press.

Köhler, Sebastian (2018). "Expressivism, Meaning, and All *That.*" *Canadian Journal of Philosophy* 48/3-4: 337-56.

Kratzer, Angelika (2012) *Modals and Conditionals.* Oxford Studies in Theoretical Linguistics. Oxford University Press.

MacFarlane, John (2004). "In What Sense (if Any) Is Logic Normative for Thought?" Unpublished.

McPherson, Tristram (2019) "Deliberative Authority and Representational Determinacy: A Challenge for Normative Realists." *Ergo* 6: 1331-58.

McPherson, Tristram (2020). "Expressivism without Minimalism." In W. Dunaway and D. Plunkett (eds.). *Meaning, Decision and Norms: Themes from the work of Allan Gibbard.* Oxford University Press.

Portner, Paul. "Imperatives." In Maria Aloni & Paul Dekker (eds.). *The Cambridge Handbook of Formal Semantics.* Cambridge. 593–628. (2016).

Plunkett, David & Timothy Sundell (2013). "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosophers' Imprint* 13/23: 1-37.

Price, Huw (1990). "Why 'Not'?" *Mind* 99 (394): 221-38.

Restall, Greg (2005). "Multiple Conclusions." In Petr Hájek, Luis Valdés-Villanueva & Dag Westerståhl (eds.). *Logic, Methodology and Philosophy of Science.* College Publications.

Ridge, Michael (2014). *Impassioned Belief.* Oxford University Press.

Rumfitt, Ian (2010). "'Yes' and 'No'." *Mind* 109/436: 781-823.

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

Schroeder, Mark (2011). "Oughts, Agents, and Actions." *The Philosophical Review* 120/1: 1-41.

Schwitzgebel, Eric (2002). "A Phenomenal, Dispositional Account of Belief." *Nous* 36 (2): 249-75.

Schwitzgebel, Eric (2013). "A Dispositional Approach to the Attitudes." In Nikolaj Nottelmann (ed.). *New Essays on Belief: Constitution, Content and Structure.* Palgrave.

Steinberger, Florian (2019). "Three Ways in Which Logic Might Be Normative." *Journal of Philosophy* 116 (1): 5-31.

Thomasson, Amie (2007). "Modal Normativism and the Methods of Metaphysics." *Philosophical Topics* 35 (1/2): 135-60.

Thomasson, Amie (2013). "Norms and Necessity." *Southern Journal of Philosophy* 51 (2): 143-60.

Thomasson, Amie (2017). "Metaphysical Disputes and Metalinguistic Negotiations." *Analytic Philosophy* 57 (4): 1-28.

Thomasson, Amie (2020) *Norms and Necessity.* Oxford University Press.

Tiefensee, Christine (forthcoming). "Metasemantics for the Relaxed." *Oxford Studies in Metaethics* 16. R. Shafer-Landau (ed.). Oxford University Press.

Unwin, Nicolas (1999). "Quasi-Realism, Negation and the Frege-Geach Problem." *Philosophical Quarterly* 49 (196): 337-52.

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

Wedgwood, Ralph (2007) *The Nature of Normativity.* Oxford University Press.

Willer, Malte (2013). "Dynamics of Epistemic Modality." *The Philosophical Review* 122/1: 45-92.

Woods, Jack (2018a). "A Commitment-Theoretic Account of Moore's Paradox." In. K.P. Turner and L. Horn (eds.). *Pragmatics, Truth and Underspecification; Towards an Atlas of Meaning.* Brill.

Woods, Jack (2018b). "The Authority of Formality." *Oxford Studies in Metaethics* 13. R. Shafer-Landau (ed.). Oxford University Press.

Yalcin, Seth (2007). "Epistemic Modals." *Mind* 116 (464):983-1026.

Yalcin, Seth (2011). "Nonfactualism about Epistemic Modals." In A. Egan and B. Weatherson (eds.). *Epistemic Modality.* Oxford University Press.

Yalcin, Seth (2014). "Semantics and metasemantics in the context of generative grammar." In Alexis Burgess Brett Sherman (eds.), *Metasemantics: New Essays on the Foundations of Meaning.* Oxford University Press. pp. 17-54.