Mark Schroeder
University of Southern California
December 1, 2011

tempered expressivism
for *Oxford Studies in Metaethics*, volume 8

This paper has two main goals. Its overarching goal, like that of some of my earlier work on expressivism, is to investigate how far one can go in developing an expressivist theory with limited resources in the philosophy of mind, and to explore the commitments one incurs in doing so. Its subsidiary goal, which will be much more in focus throughout, is to introduce what I take to be a new kind of expressivist theory – one which goes particularly far with particularly limited resources from the philosophy of mind. This new kind of theory is not, as you might expect, called 'tempered expressivism', but 'relational expressivism'. But the most important concept to understand, in order to see the main ideas behind relational expressivism, is that of a tempered expressivist theory.

1.1   expressivism: tempered and untempered

The basic idea of expressivism is that for some sentences 'P', believing that P is not just a matter of having an ordinary descriptive belief. This is a way of capturing the idea that the meaning of some sentences either exceeds their factual/descriptive content or doesn't consist in any particular factual/descriptive content at all, even in context. The paradigmatic application for expressivism is within metaethics, and holds that believing that stealing is wrong involves having some kind of desire-like attitude, with world-to-mind direction of fit, either in place of, or in addition to, being in a representational state of mind with mind-to-world direction of fit. Because expressivists refer to the state of believing that P as the state of mind 'expressed' by 'P', this view can also be described as the view that 'stealing is wrong' expresses a state of mind that involves a desire-like attitude instead of, or in addition to, a representational state of mind.

According to some expressivists – *unrestrained* expressivists, as I'll call them – there need be no special relationship among the different kinds of state of mind that can be expressed by sentences. Pick your favorite state of mind, the unrestrained expressivist allows, and there could, at least in principle, be a sentence that expressed it. Expressivists who seem to have been unrestrained plausibly include Ayer in

Language, Truth, and Logic, and Simon Blackburn in many of his writings, including his [1984], [1993], and [1998].

But there are a number of reasons for expressivists not to be unrestrained in the freedom that they allow for sentences to express arbitrary states of mind. For example, it doesn't seem possible to disagree with just any state of mind, yet declarative sentences all seem to express states of mind with which it is possible to disagree. This leads Allan Gibbard [2003] to restrain his expressivism with the restriction that only states of mind with which it is possible to disagree may be expressed by declarative sentences. Terry Horgan and Mark Timmons [2006] also restrain their expressivism with a restriction on which sorts of states of mind can be expressed by declarative sentences to those which share a common set of rational and phenomenological properties characteristic of beliefs.

A different reason to restrain one's claims about what kinds of state of mind can be expressed by declarative sentences derives from the famous Frege-Geach Problem for expressivism. In general, the Frege-Geach Problem is the problem of saying what state of mind is expressed by complex sentences. But some complex sentences are mixed. For example, consider the disjunction, 'either the economy will soon recover on its own, or the Fed ought to pursue a more expansionary policy'. This is a disjunction with one normative disjunct and one non-normative disjunct. If the states of mind expressed by such disjuncts could bear no relation to each other whatsoever, as the wholly untempered expressivist allows, it's quite puzzling what the state of mind expressed by the disjunction is. It looks like it can be neither of these two kinds of state of mind, but must bear some important relationship to each, because it constrains the ways in which they can be rationally combined. According to biforcated attitude semantics as developed in Schroeder [2008], though the states of mind expressed by normative and non-normative sentences can be quite different from one another, every state of mind expressible by a declarative sentence must belong to a common kind – a 'biforcated attitude'.[1] This is the key, in biforcated attitude semantics, to making progress with the Frege-Geach Problem.

So Gibbard [2003], Horgan and Timmons [2006], and Schroeder [2008] all describe restrained versions of expressivism. Yet on each of these forms of expressivism, though there are some restrictions on what sort of state can be expressed by sentences, the states of mind expressed can still be quite different, and there is no privileged role for paradigmatic or, as I will call it, *ordinary descriptive* belief to play. On each of these pictures, the kind of ordinary descriptive belief expressed by 'grass is green' is just one special case

---

[1] 'Biforcated' is a pun, not a typo; such states consist in two states of 'being for'. See Schroeder [2008].

among others of the kind of state of mind expressible by a declarative sentence. I'll call such views *restrained* but *untempered*.

In contrast to both the unrestrained views of Ayer and Blackburn and the restrained but untempered views of Gibbard and Horgan and Timmons, the *tempered* expressivist holds that there is an asymmetric relationship between the kind of ordinary, factual/descriptive representational state of mind expressed by an ordinary sentence like 'grass is green' and the kinds of state of mind that can be expressed by other sentences. The tempered expressivist holds that in order to be expressible by a declarative sentence, a state of mind needs to be *belief-involving*. That is, it needs to 'involve' ordinary descriptive belief, in some way. So for the tempered expressivist, ordinary descriptive sentences have a kind of priority and centrality that they lack on untempered views. Tempering is therefore a very special kind of restraint.

When I say that the states of mind expressible by declarative sentences must be 'belief-involving', I mean that they must involve the kind of state of mind which *cognitivists* hold moral beliefs to be – the kind of state that I've been calling *ordinary descriptive belief*. There is a less substantive way in which all expressivists may say that the states of mind expressible by declarative sentences involve belief – because *all* declarative sentences express beliefs. On this view, what is distinctive of metaethical expressivism is the view that moral beliefs are different in kind from non-moral beliefs. An expressivist who says that all declarative sentences express beliefs (as I believe she should) can re-frame the issues surrounding restraint and tempering as not about what constraints there are on what sort of state can be expressed (trivially: it must be a belief), but rather about what characteristics a state of mind must satisfy in order to be a belief. On this characterization, all forms of restraint – tempering included – can be thought of as substantive hypotheses about the nature of belief. Unrestrained views, so construed, amount to the claim that belief has no essential nature, so that ordinary, non-philosophical talk of 'belief' is purely deflationary.[2]

So far my characterization of tempering is highly abstract. It says only that the tempered expressivist must hold that the states expressible by declarative sentences must 'involve' ordinary descriptive belief, but I haven't said anything to make this 'involvement' thesis precise. It turns out that there are at least two important ways to make good on this thesis, leading to two different forms of tempered expressivism, one very familiar from the literature, and one, I believe, much less familiar, but equally worthy of attention.

---

[2] Compare Ridge [2007a] on 'minimalist' belief, and contrast in particular Horgan and Timmons [2006].

1.2   hybrid expressivism

The most familiar form of tempered expressivism is what I've elsewhere called hybrid expressivism.[3] The basic idea of hybrid metaethical expressivism is that moral sentences express states of mind that consist in both an ordinary descriptive belief and a desire-like attitude. A helpful comparison is to an intuitive model for slurring language. On this model, to accept a sentence[4] like 'Schroeder is a cheesehead' your state of mind needs to satisfy two conditions: you must both believe that Schroeder is from Wisconsin, and also have some sort of contemptuous attitude toward people from Wisconsin. The core hybrid idea is that something like this might be what is going on with normative language more generally, including 'thin' evaluatives like 'good' and 'wrong'.

On this picture, there is some particular property – call it K – such that you can't believe that stealing is wrong without believing that stealing instantiates K, and there is, in addition, some desire-like attitude with respect to K, such that you can't believe that stealing is wrong unless you have that attitude. For example, just to put a less abstract face on the hybrid view, a particular hybrid theorist might fill in this schematic picture by holding that 'stealing is wrong' expresses a state of mind which combines the belief that stealing fails to maximize happiness with the desire to avoid doing what fails to maximize happiness. So believing that stealing is wrong consists in having each of these two attitudes. This hybrid view can explain why anyone who believes that stealing is wrong will be motivated not to steal, because on this view any such person will believe that stealing has a property that she desires to avoid. So views with this structure share with non-hybrid expressivist views the capacity to explain why moral judgments motivate.

Theorists who have adopted something along the lines of the hybrid view I've just described include David Copp [2001], [2009], Daniel Boisvert [2007], Ryan Hay [forthcoming], Stephen Barker [2000], and on a literal reading, Michael Ridge [2006], [2007a], [2007b]. The most prominent difference between the views of Copp, Boisvert, and Hay, on the one hand, and those of Barker and Ridge, on the other, is that Barker and Ridge hold that both components of the state of mind expressed by moral sentences vary from speaker to speaker as a result of the contextual contribution of the moral terms, whereas for Copp, Boisvert, and Hay, this contribution is context-invariant, or at least does not vary purely as a result of the speaker. In Schroeder [2009] I offered arguments against the Barker/Ridge way of

---

[3] See Schroeder [2009].

[4] I will sometimes use 'accept the sentence' as a shorthand for 'be in the mental state expressed by'. Recall that intuitively, being in the mental state expressed by 'P' is just believing that P. But some sentences 'P' contain terminology that we as theorists might not wish to use – for example, 'cheesehead'. So as theorists we should prefer not to need to describe the meaning of 'Schroeder is a cheesehead' by saying what it is to believe that Schroeder is a cheesehead. This makes it convenient to take the expression relation as relatively primitive, rather than defined in terms of belief, and to cast the gloss in terms of belief as just a gloss. See Schroeder [forthcoming] for further discussion of this refinement.

developing a hybrid theory and for purposes of this paper I'll assume the context-invariant version of the hybrid theory for ease of discussion. (I'll return to this issue in section 4.2.) Moreover, although there are many ways of hybridizing a metaethical theory to include both descriptive and noncognitivist components, I will focus in what follows on a version of this strategy that works within a traditional expressivist framework, simply by adopting the view that the state of mind expressed by normative sentences has an ordinary belief component, and I will focus on versions which obtain the 'virtues' mentioned in the title of Schroeder [2009].[5]

It is important to note that it is perfectly intelligible to hold that something like this picture is correct for some terms – perhaps including pejoratives, slurs, and some other evaluative language – while holding that there are other sentences that express totally different kinds of state which needn't involve ordinary descriptive belief in any way. For example, you could be a hybrid theorist about 'wrong' and 'cheesehead' but a pure expressivist about 'ought' and 'good'.[6] So it is perfectly possible to be a hybrid expressivist about some terms in the language while being an untempered or even unrestrained expressivist in general.

But it is particularly natural for a hybrid expressivist to be a tempered expressivist, because one of the prominent motivations offered by hybrid theorists for their views[7] has been that it is easier to see how sentences that express hybrid states of mind can get around the well-known Frege-Geach or 'embedding' problem for expressivism, and that there are ample grounds for optimism that this project can be successful, because of the example of slur terms, which embed in all of the same ways that moral terms do more generally, but plausibly appear to have a hybrid character. But the example of slurs can make us optimistic only if we think that everything works that way – if there are some sentences that work on a different model, the example of slurs gives no particular grounds for optimism that they will not face an intransigent embedding problem. So any hybrid theorist who takes refuge in the precedent of slurs as grounds for optimism must implicitly be assuming not only that moral sentences express hybrid states of mind, but that no declarative sentences express states of mind that do not involve belief.

---

[5] It is important to note that there are many (actual and possible) hybrid metaethical theories which do not operate within the expressivist paradigm, and that many hybrid expressivist theorists describe their view as that moral sentences express each of two states of mind, rather than (as I have described the view) that they express a single state of mind that wholly consists in the conjunction of a belief state and an attitudinal state. I don't believe that anything important turns on the question of which of these ways to characterize the view.

[6] Note that any semantic theory, expressivist or descriptivist, can be hybridized by the same methods that are used to hybridize descriptivist theories, as discussed in Schroeder [2009]. So to be an untempered hybrid expressivist, you need only to add a hybrid component to an existing expressivist view. It is even possible to defend a hybrid theory that is tempered, but not in the way distinctive of tempered hybrid theories, by starting with a tempered *relational* theory, as described below, and then adding a distinct hybrid component for slurs and other pejoratives.

[7] Compare especially Copp [2001], Boisvert [2007], Ridge [2006], [2007a].

A hybrid expressivist who takes such a view is *tempered*, in the sense outlined above. If every declarative sentence expresses a state of mind that consists in part in a particular ordinary descriptive belief, then that makes ordinary descriptive sentences, which *just* express ordinary descriptive beliefs, out to be a privileged and central case. As I've just noted, the example of slurs provides license for optimism that there will be no insurmountable problems for such a view, and in part 2 of this paper I will explain in more detail how this kind of tempered hybrid theory in fact avoids some of the most striking challenges to traditional 'pure' expressivist theories. But first I need to introduce a second, less familiar, way in which an expressivist theory may be tempered.

1.3 relational expressivism

Despite its familiarity, hybrid expressivism is only one way of tempering expressivism. A second, less familiar, way of tempering expressivism is offered by a kind of view that I'll call *relational* expressivism. Although there are important antecedents and possible instances of it in the literature, as I'll discuss in part 3, I know of no clear and explicit articulations of this view in the literature to date.[8] But as we'll see, it is an important alternative to hybrid expressivism, sharing many of its virtues but coming with a very different set of commitments.

The basic idea of the relational expressivist is that moral beliefs consist in a certain *relation* holding between one's ordinary descriptive belief state and some kind of desire-like attitudinal state. Like the hybrid theorist, the relational expressivist holds that moral beliefs are nothing over and above ordinary descriptive beliefs together with desire-like attitudes. But whereas the hybrid theorist holds that there is a particular pair of an ordinary descriptive belief and a desire-like attitude that you must have, in order to have a moral belief, the relational expressivist holds only that a certain relation must obtain between your ordinary descriptive beliefs and your desire-like attitudes.

It is important at the outset to make two clarifications about what sort of relation the relational expressivist may allow to be involved in the state expressed by a declarative sentence. Since some non-moral sentences surely express states of mind that it is possible to realize wholly in virtue of one's ordinary descriptive belief state, the relational expressivist must allow that these relations may be vacuous in their attitudinal relatum.[9] A relation between one's ordinary descriptive belief state and one's attitudinal state that is vacuous in the second argument is just a property of one's ordinary descriptive belief state.

---

[8] However, see Toppinen [this volume].
[9] Here and in what follows for readability I will sometimes drop the qualifier 'desire-like' on 'attitude' or 'attitudinal'. Though strictly speaking beliefs are a kind of attitude, I will throughout use 'attitude' for non-cognitive states.

But importantly, in order to be tempered, the relational expressivist must rule out the possibility that this relation can be vacuous in the ordinary descriptive belief relatum. For if it were, then such a relational state of mind could be realized by an agent irrespective of her ordinary descriptive belief state, and hence it would not 'involve' ordinary descriptive belief in any natural sense of 'involve'. So the tempered relational expressivist will hold that all declarative sentences express states of mind consisting in a relation between one's ordinary descriptive belief state and one's attitudinal state that may be vacuous in the second relatum but not in the first.

If tempered expressivism sounded abstract to you to begin with, this gloss on relational expressivism is clearly more so. So it will be helpful to get a grip on what a particular relational expressivist theory might look like, and a sense of what independent motivation it might have. I'll do so by analogy. In *Thinking How to Live*, Allan Gibbard described human psychology as involving at least two important kinds of psychological states: prosaic beliefs, and conditional plans. Prosaic belief states are what I'm here calling 'ordinary descriptive beliefs, and they play the role for the agent of keeping track of which circumstances she finds herself in. Conditional plans are plans for what to do in any of a wide range of possible situations the agent might find herself to be in – for Gibbard, even situations it would be impossible for her to get into.

One might think that in addition to prosaic beliefs and conditional plans, agents must also have unconditional or categorical plans. But this is far from obvious. Once agents have conditional plans about what to do in a wide variety of circumstances and beliefs about which circumstances she is actually in, that should settle what her categorical plan is: it should be to do whatever she conditionally plans to do, conditional on finding herself in the situation she believes herself to be in. So on one natural idea, this is all that categorical plans are – a relationship between your conditional plans and your prosaic beliefs. On this view, to plan to do something is just to conditionally plan to do it, in the situation you believe yourself to be in.

An intuitive way of seeing how disapproval – the state of mind expressed by 'wrong' sentences – could be a relational state of mind is to follow this model. Suppose that each agent has a prosaic belief state and a conditional disapproval state – an attitude toward proposition-action pairs which tells us which actions the agent disapproves of conditional on that proposition's being the case. Then we can think of what she categorically disapproves of as simply what she conditionally disapproves of, conditional on the totality of her beliefs. This gives a picture – just one picture, but one that I think can be initially illuminating – of what relational states of mind like those posited by relational expressivism could be like, and of why they are worth taking seriously.

7

So on this view, believing that something is wrong consists in a relation between your conditional disapproval state and your ordinary descriptive belief state. So there is a natural sense in which it 'involves' ordinary descriptive belief, but in a different way from the hybrid view. On the hybrid view, believing that stealing is wrong involved your ordinary descriptive belief state by placing a particular constraint on which ordinary descriptive belief you need to have, at least holding context fixed. But on this relational view, there is no particular ordinary descriptive belief that you need to have – just a relation to your conditional disapproval state. Similarly, on the hybrid view, believing that stealing is wrong involves having a particular desire-like attitude, at least holding context fixed. But on the relational view, in contrast, there is no particular conditional disapproval state that you need to be in (so long as you don't tolerate it unconditionally); there only needs to be the right relation between your conditional disapproval state and your ordinary descriptive belief state.

It's easy to see that although I've described relational expressivism as assuming an underlying state of mind that I've called 'conditional disapproval', there will be a range of alternative pictures of what this attitude is like. What is important to the view is that moral beliefs consist in a relation between one's ordinary descriptive belief state and some desire-like attitude state. It is to some degree natural, for purposes of making the view intuitive, to think of this as a kind of conditional attitude state, and for convenience of exposition I will stick to glossing it as conditional disapproval. But it is important to keep in mind that this is for illustrative purposes only.

As with hybrid expressivism, it is important to note that nothing about relational expressivism as I've described it so far dictates that it amounts to a kind of tempered expressivism. It is perfectly intelligible to hold that moral sentences express such relational states of mind, but allow, contrary to tempering, that other sentences express non-relational states of mind that have nothing to do with ordinary descriptive belief, or even allow, contrary to restraint, that declarative sentences may express any kind of state of mind at all.

But as we'll see in what follows, it is natural for a relational expressivist to hope that by tempering her theory, she can use the features of ordinary descriptive belief in order to help her avoid some of the important challenges for other expressivist views. A relational expressivist view is tempered if it holds that *all* declarative sentences must express states of mind that consist in a relation between one's state of conditional disapproval and one's ordinary descriptive belief state that is non-vacuous in the ordinary descriptive belief relatum.

In this and the last section I've introduced hybrid expressivism and relational expressivism as alternative ways of developing the tempered approach to expressivism. The promise of such theories is

closely related to their prospects to dissolve some of the chief challenges facing other versions of expressivism. In part 2 I'll shed more light on how both hybrid expressivism and relational expressivism work, by illustrating their resources for dealing with two of the most striking challenges for other forms of metaethical expressivism: the Frege-Geach or 'embedding' problem, and Cian Dorr's wishful thinking problem. Once their virtues are on the table, we'll turn in part 3 to compare them to each other and explore the nature and prospects of their commitments.

2.1   the frege-geach problem

The most famous obstacle confronting traditional forms of expressivism is what has come to be known as the 'Frege-Geach' or 'embedding' problem.[10] The essence of the problem is that it is not enough, to characterize the meaning of a moral word, to say what state of mind is involved in accepting simple sentences involving that word. You must also explain how such simple sentences combine to yield the meaning of complex sentences. For example, you must be able to say what it is to believe that stealing is not wrong, to believe that either stealing is wrong or it doesn't hurt anybody, and that if stealing is wrong, then so is killing.

There are two main parts to the difficulties of traditional expressivist and other noncognitivist theories in dealing with this obstacle. The first part consists in the fact that it is hard to see what account the expressivist can give of the thoughts corresponding to complex sentences at all. The second part consists in the challenge, even after one has such an account, of explaining why the those thoughts have the right properties. I'll take each of these aspects of the challenge in turn, and further divide each into two illustrative component challenges.

Traditional theories of meaning are broadly truth-conditional, in the sense that they account for the meaning of sentences by characterizing the conditions under which they are true, or by assigning propositions which fix those conditions. Truth conditions obey important principles which facilitate using them for compositional semantics. Importantly, the conditions under which '~P' is true are just those under which 'P' is not true, the conditions under which 'P&Q' is true are just those under which 'P' is true and 'Q' is true, and the conditions under which 'P∨Q' is true are just those under which 'P' is true or 'Q' is true. I call these the properties of *complementation*, *intersection*, and *union*, respectively, because when truth conditions are represented as sets of circumstances under which the sentence is true, they can be characterized in terms of the set-theoretic operations of complementation, intersection, and union. The

---

[10] See especially Searle [1962], Geach [1965], van Roojen [1996], Unwin [1999], and Schroeder [2010]

fact that truth-conditions obey these principles is an important part of the success story of truth-conditional compositional semantics, because they are what make it possible to read the truth conditions of complex sentences off of the truth conditions of their parts.

In contrast, the semantic values in which the expressivist is primarily interested – states of mind – don't obey these principles. Believing that ~P is very different from not believing that P, believing that P&Q is different from both believing that P and believing that Q, and believing that P∨Q is very different from either believing that P or believing that Q. This is, at bottom, why broadly truth-conditional compositional methods are inappropriate for expressivism. Expressivists need a different sort of account of how to generate the meanings of complex sentences from the meanings of their parts.

More strongly, in contrast to traditional theories of meaning, which require no special theory about the mind, expressivists need a special account of the mind in order to give their compositional semantics – an account of the mind that goes over and above anything that they have to say in order to give their account of simple sentences. One way that we can illustrate this point is with the case of mixed normative-descriptive disjunctions, like 'either stealing is wrong or it doesn't hurt anyone'. In order to give their compositional semantics, expressivists must tell us what it is to believe that either stealing is wrong or it doesn't hurt anyone.

It is an important and striking feature of this state of mind that it is related to the kind of state of mind expressed by 'stealing is wrong' and to the kind of state of mind expressed by 'it doesn't hurt anyone'. In particular, together with the state of mind expressed by 'stealing is not wrong', it commits a thinker to the state of mind expressed by 'it doesn't hurt anyone', and together with the state of mind expressed by 'it hurts someone', it commits a thinker to the state of mind expressed by 'stealing is wrong'. But according to the expressivist, the state of mind expressed by 'stealing is wrong' is importantly different in kind from the state of mind expressed by 'it doesn't hurt anyone'. If these are such different states of mind, it is intuitively hard to see how the disjunction could express either, because it needs to constrain their relationship.

But the disjunctive belief can't *merely* constrain the relationship between these two kinds of state of mind, either. For one state of mind that merely constrains this relationship is the state of either believing that stealing is wrong or believing that it doesn't hurt anyone. This state has all of the inferential relationships that the disjunctive belief has, but it's not the same as the disjunctive belief, because it's possible to believe the disjunction without believing either disjunct. In order to allow for this, the expressivist must tell us what state of mind is involved in believing the disjunction without believing either

disjunct. (Note that this is just the task of explaining why belief does not obey the principle of union!) It turns out that this is difficult to do, and expressivist theories have not always managed to clear this minimal bar.[11]

So far I've been using the example of disjunctive belief in order to illustrate why there is a special challenge to expressivists, at least in general, to even tell us what state of mind is expressed by complex sentences that even succeeds at making elementary distinctions, such as the distinction between believing that ~P and not believing that P, and the distinction between believing that P∨Q and either believing that P or believing that Q. This general problem is illustrated by two component challenges: the *productivity* challenge of being able to provide a general compositional principle which gives any answer at all to what mental state is expressed by complex sentences, and the *conflation* challenge to avoid making inappropriate conflations, such as conflating the state of believing a disjunction with the state of believing one of the disjuncts.[12]

But even once expressivists have an account that can make these distinctions, they need to be able to use that account in order to explain why complex sentences have the right semantic properties. This is the second important aspect of the Frege-Geach problem. For example, the sentence 'stealing is not wrong' is inconsistent with the sentence 'stealing is wrong', and that is not a coincidence; it is due to the meaning of the word 'not'. So whatever compositional principle expressivists supply for the meaning of 'not' needs to be able to predict and explain why this should be so. Similarly, since conditionals validate *modus ponens*, any expressivist compositional principle for the conditional must be able to explain why this is so. This is essentially the challenge that Geach [1965] laid down, on a charitable reading.[13]

It turns out that expressivist explanations of inconsistency and validity are somewhat complicated, because inconsistency and validity are closely related to truth, and the expressivist's view about truth takes some work to develop.[14] But even setting truth aside, there are more basic features of valid arguments that an expressivist should be able to explain. In particular, an expressivist compositional rule for the conditional should be able to explain why it is rationally inconsistent to accept the premises of a *modus ponens* argument while denying its conclusion – an important property of classically valid arguments that I call the *inconsistency property*. And it should also be able to explain why accepting the premises of a *modus*

---

[11] Just to take one recent example, the expressivist view of epistemic modals elaborated in Yalcin [2007] predicts that all three of the principles of complementation, intersection, and union hold whenever there are terms involved that require an expressivist treatment. See Schroeder [unpublished] for discussion.

[12] See especially Unwin [1999], Dreier [2006], and Schroeder [2008] for discussion of the tension between *productivity* and avoiding *conflation* for the special and illustrative case of negation.

[13] See Schroeder [2010] for further discussion.

[14] See Schroeder [2010] (chapter 8), [forthcoming] for elaboration.

ponens argument *rationally commits* you to accepting its conclusion – another important property of many (though at least arguably not all) classically valid arguments that I call the *inference-licensing property*.[15]

In section 2.2, I'll show how a tempered hybrid theory can give straightforward answers to what mental state is expressed by complex sentences, and explain why these straightforward answers yield straightforward explanations of why *modus ponens* and disjunctive syllogism satisfy both the inconsistency property and the inference-licensing property. Then in section 2.3 I'll provide a similar picture for the relational expressivist. Finally, in section 2.4 I'll introduce Cian Dorr's wishful thinking problem and show how both kinds of tempered view deal with it.

2.2   the hybrid solution

The tempered hybrid expressivist's solution to the Frege-Geach problem is closely modeled on observations about its model case from natural language: slurs. Plausibly, if accepting 'Schroeder is a cheesehead' requires having a contemptuous attitude toward people from Wisconsin, accepting 'Schroeder is not a cheesehead' also requires having such an attitude. For those of us who don't have such an attitude toward people from Wisconsin, 'cheesehead' is, as we might put it, *not one of our words*, and we needn't accept any sentences involving it. This is not the only natural view to take about slurs like 'cheesehead', but it is one common and natural picture, and it is the picture that motivates the hybrid theorist's compositional semantics. The picture gives us an easy answer as to how 'cheesehead' works compositionally: every complex sentence containing 'cheesehead' expresses a state of mind that involves the same contemptuous attitude as simple sentences containing 'cheesehead' do, along with an ordinary descriptive belief whose content is composed in the normal way.[16]

The hybrid expressivist extends this same model to moral sentences. According to this picture, each normative word has both an ordinary descriptive content and an associated desire-like attitude. The mental state expressed by an arbitrary sentence consists in a belief component, whose content is determined compositionally in the normal way, together with each desire-like attitude that is associated with any of its component words. More precisely, on this picture, if 'P' expresses the state consisting of the belief that P^* and attitudes $A_{p1} \ldots A_{pn}$, and 'Q' expresses the state consisting of the belief that Q^* and attitudes $A_{q1} \ldots A_{qm}$, then '$\sim$P' expresses the state consisting of the belief that $\sim P^*$ and attitudes $A_{p1} \ldots A_{pn}$, 'P&Q' expresses the state consisting of the belief that $P^* \& Q^*$, attitudes $A_{p1} \ldots A_{pn}$, and attitudes $A_{q1} \ldots A_{qm}$, and 'P$\vee$Q' expresses

---

the state consisting of the belief that $P^* \lor Q^*$, attitudes $A_{p1}...A_{pn}$, and attitudes $A_{q1}...A_{qm}$. On this picture, the two components of the attitude expressed by normative senences – the belief and the desire-like attitudes – make different contributions to compositional semantics. The belief contributes its content to a standard compositional picture, and the desire-like attitude simply gets 'kicked up' to complex sentences in which it figures.[17]

There is no puzzle, on this view, as to what state of mind is expressed by sentences like 'either stealing is wrong or it doesn't hurt anyone'. It consists in the combination of an ordinary descriptive belief to the effect that either stealing instantiates K or it doesn't hurt anyone, together with the desire not to do what instantiates K. It is also easy to see that it is possible, on this view, to accept this disjunction without accepting either disjunct. All that you have to do is to desire not to do what instantiates K and have the corresponding ordinary descriptive belief without believing either disjunct. Since it's possible to have ordinary descriptive disjunctive beliefs without believing either disjunct, it's possible to have normative disjunctive beliefs without believing either disjunct.

It's also straightforward on this view why all classically valid arguments have the inconsistency property. For the hybrid compositional semantics guarantees that there is a structure-preserving isomorphism between sentences and the ordinary descriptive beliefs involved in the states of mind that they express. So accepting the premises of any classically valid argument while denying its conclusion involves having a set of ordinary descriptive beliefs whose contents are classically inconsistent. So whatever explains why ordinary descriptive classically valid arguments have the inconsistency property will suffice to explain why all classically valid arguments do.[18]

And finally, it is also straightforward how the hybrid theorist accounts for the inference-licensing property of important classical inferences like *modus ponens*, disjunctive syllogism, and conjunction introduction. Because of the structural isomorphism between sentences and the descriptive belief component of the states of mind they express, accepting the premises of a valid argument involves having ordinary descriptive beliefs which rationally commit you to having the ordinary descriptive belief that is involved in accepting the conclusion.

This doesn't *quite* commit you to accepting the conclusion in the case of *every* classically valid argument, because whenever there are normative words in the conclusion, it takes more to accept it than to

---

[17] It turns out that there are important reasons why a hybrid theorist might not wish to extend this 'kicking up' story to embedding under attitude verbs, but I won't be concerned with those reasons, here. See Schroeder [2009] (sections 8 and 9) and Hay [2011] for discussion. See also Potts [2005] for the extensive development of a structurally similar view in a non-expressivist framework.

[18] Compare Ridge [2004], Boisvert [2007], Schroeder [2009], and Hay [2011].

have the associated ordinary descriptive belief – you must also have each of the desire-like attitudes associated with each of its normative words. But so long as every normative word that appears in the conclusion also appears in one of the premises, you *already* have these attitudes in virtue of accepting the premises. So every classically valid argument in which all of the normative words appearing in the conclusion appear in one of the premises has the inference-licensing property. This includes core productive inferences like *modus ponens*, disjunctive syllogism, and conjunction introduction.[19]

So in particular, the hybrid theorist doesn't explain why vacuous inferences to classical theorems have the inference-licensing property, which means that it doesn't predict that everyone is rationally committed to accepting each instance of classical theorems like 'P∨~P'. But this should be no surprise, given that the hybrid theorist has modeled her view on the case of slurs like 'cheesehead'. For given the view adopted about 'cheesehead', to accept 'either Schroeder is a cheesehead or he isn't a cheesehead', you need to have the contemptuous attitude toward people from Wisconsin associated with 'cheesehead'. If you don't have that attitude, you won't even accept classical theorems involving the word – it won't be one of your words. Similarly, on the hybrid view, if you don't have the right desire-like attitudes, normative words in general won't be among your words.

So though the hybrid view doesn't predict that all classically valid arguments have the inference-licensing property, it does predict this for the core inferences that don't allow vacuous introduction of new words in the conclusion, and it has a plausible explanation of why it is right not to go further. Together with the other features of the view that we've already discussed, this shows that hybrid views have straightforward, intelligible answers to each of the major aspects of the Frege-Geach problem that we've discussed: productivity, conflation, inconsistency, and inference-licensing. Since all of these answers turn crucially on the fact that all sentences express states of mind that involve ordinary descriptive belief, and because each of these answers piggy-backs on the fact that ordinary descriptive belief has all of these properties, they are the fruits of hybrid tempering. In the next section we'll turn to look how a tempered relational expressivism can deal with the same issues.

2.3   the relational solution

In the previous section I drew on the discussions of existing hybrid expressivist views in the literature in order to be able to begin by explaining how a hybrid compositional semantics works, and then show how it solves the problems we've been focusing on. In this section, in contrast, since there don't seem to be any

---

[19] See Schroeder [2009] for extended discussion of this point.

totally clear and explicitly self-aware relational expressivist views in the literature, I'll need to adopt a different strategy. The key insight we'll need is that hybrid expressivism is really just a special case of relational expressivism.

According to hybrid expressivism, the state of mind you need to be in to accept a moral sentence wholly consists in your ordinary descriptive belief state satisfying a certain condition, and your attitudinal state satisfying a certain condition. This is, in fact, a special case of a relation holding between your ordinary descriptive belief state and your attitudinal state – one that can be wholly decomposed into the conjunction of separate properties for each relatum. The relational expressivist's picture is exactly like this, but simply more general. Rather than associating each sentence with a relation between one's ordinary descriptive state and one's attitudinal state that can be wholly decomposed into properties of each relatum, the relational expressivist simply allows that the specified relation can be genuinely *relational*. With this observation in hand, the path forward for relational expressivism is clear: it should seek to generalize on the compositional methods employed by the hybrid expressivist.

It turns out, however, that not just *any* relational view can generalize on the compositional methods employed by the hybrid expressivist. Because the hybrid theorist aims to piggy-back on standard compositional methods, she doesn't allow that declarative sentences to express states that impose just *any* constraint on one's ordinary descriptive belief state. She limits her view to states that consist in having a *particular* ordinary descriptive belief, with a particular ordinary descriptive content. So if the relational expressivist hopes to piggy-back on ordinary compositional methods in the same way, then she needs to adopt the same restriction. Rather than allowing just *any* relation between one's ordinary descriptive belief state, holistically characterized, and one's attitudinal state to count, she needs to specify that only relations which depend only on whether one has a particular ordinary descriptive belief are to count.

Such relations satisfy the important property of being *belief-monotonic*, in the sense that if they hold between an attitudinal state and any ordinary descriptive belief state, then they also hold between that same attitudinal state and any strictly more opinionated ordinary descriptive belief state. It's a trivial observation that the standard hybrid compositional semantic strategy that I've described also requires the relational state expressed by a sentence to be belief-monotonic.

It is important to note that the requirement of belief-monotonicity is far from trivial. In section 1.3, when I tried to make my characterization of relational expressivism less abstract, I offered a comparison to the idea that what an agent actually plans to do (unconditionally) is just a relation between what she *conditionally* plans to do in different circumstances, and the circumstances she believes herself to be in. But planning is clearly *not* belief-monotonic. Sophia may plan to go to school today because she

15

conditionally plans to go to school on the condition that it's a weekday and believes that it's a weekday. But it doesn't follow that she conditionally plans to go to school on the condition that it's a weekday and her grandmother's funeral is scheduled. So because conditional plans are nonmonotonic, the relation between ordinary descriptive beliefs and conditional plans that constitutes outright plans fails to be belief-monotonic. So the restriction to belief-monotonic relations is non-trivial.

With this restriction on board, however, the relational expressivist can fully characterize her view by assigning each declarative sentence 'P' to a function from attitudinal states to propositional contents. If A is the set of all possible attitudinal states and C is the set of propositional contents, then we may say, for any sentence 'P' that $[P]:A{\to}C$ is a function mapping each attitudinal state α in A to the proposition such that it is necessary and sufficient for someone who is in α to believe, in order to count as believing that P. This notation makes it simple to state the compositional rules. For all attitudinal states α, $[{\sim}P]\alpha$ is the negation of $[P]\alpha$, $[P\&Q]\alpha$ is the conjunction of $[P]\alpha$ and $[Q]\alpha$, and $[P{\vee}Q]\alpha$ is the disjunction of $[P]\alpha$ and $[Q]\alpha$.

It is easy to observe that the hybrid compositional semantics is a special case of these rules – the special case we get when we stipulate that for all 'P', $[P]$ is a constant-valued function whereever it is defined. When $[P]$ is defined for every attitude, this tells us that there is a single proposition such that no matter what your attitudinal state, believing it is necessary (and sufficient) for believing that P. This is obviously the ordinary descriptive case. And when $[P]$ is undefined for some attitudinal states, that tells us that it is impossible to believe that P unless your attitudinal state is one of the ones for which it is defined – in which case again there is a single proposition belief in which is necessary (and sufficient, given your attitudinal state) for believing that P. So these compositional rules are a strict generalization on the hybrid compositional picture.

Finally, with this picture on board, it is straightforward to observe how the relational expressivist can piggy-back on the hybrid solution to each of the four main aspects of the Frege-Geach problem that I've noted. First, it is clearly *productive*, and in the same way as any other account which assigns declarative sentences to ordinary descriptive contents. And it avoids conflation in exactly the same way. No matter what your attitudinal state, and for any sentences 'P' and 'Q', there is some disjunctive ordinary descriptive proposition $p{\vee}q$ such that it is necessary and sufficient for you to believe that P∨Q, that you believe $p{\vee}q$. But since it is necessary for you to believe that P, that you believe p, and necessary for you to believe Q that you believe q, all we need in order to explain why it is possible to believe that P∨Q without believing that P or believing that Q, is that ordinary descriptive belief avoids exactly this conflation, so that it is possible to

16

believe $p \lor q$ without believing p or believing q. So again, we get to piggy-back on the fact that ordinary descriptive belief has these properties, and in the same way as for hybrid expressivism.

It's also easy to see that the relational expressivist can piggy-back on the hybrid theorist's explanations of the inconsistency and inference-licensing properties. For holding fixed your conditional disapproval state, the compositional semantics induces a structure-preserving isomorphism between sentences and the ordinary descriptive beliefs that you need to have, in order to accept those sentences. So if you accept the premises of a classically valid argument and deny its conclusion, then you have ordinary descriptive beliefs with classically inconsistent contents, getting us the inconsistency property.

Similarly, if you accept the premises of a classically valid argument, you have ordinary descriptive beliefs which commit you to the ordinary descriptive belief that you need to have in order to accept its conclusion. In contrast to the hybrid theorist's explanation of the inference-licensing property, however, which only explained why *some* classically valid arguments have the inference-licensing property, the relational expressivist is – at least in principle – able to get an explanation of why *any* classically valid argument does. This is because no matter what your attitudinal state is, in accepting the premises you have ordinary descriptive beliefs that commit you to the ordinary descriptive belief that is sufficient for accepting the conclusion given your *actual* attitudinal state.[20]

In this section I've shown how starting with the idea of piggy-backing on the way in which the hybrid theorist avoids conflating belief in a disjunction with belief in one of the disjuncts, the relational strategy can co-opt the hybrid theorist's methods more generally to avoid each of the difficulties associated with the Frege-Geach problem on which we've been focusing: with composition, conflation, inconsistency, and inference-licensing. Indeed, its account of inference-licensing has the potential to be more comprehensive than that of the hybrid theorist.

Each of these answers depends on the assumption that the mental state associated with every sentence involves a relation to an ordinary descriptive belief state. Because the semantics operates on the belief states, this relation must be non-vacuous in the belief relatum, but because it only requires that we be able to perform the same operation for each possible conditional disapproval state, it leaves open that this relation could be vacuous in its other relatum, relating every conditional disapproval state to the same ordinary descriptive belief state. So this shows the fruits of relational tempering for addressing the central difficulties associated with the Frege-Geach Problem.

---

[20] This result turns on requiring that for all 'P', [P] is defined throughout A. As we saw, when we conceive of the hybrid view as a special case of the relational view, it assigns sentences to functions that are undefined for some attitudinal states. So no hybrid view is a 'special case' of a relational theory that adopts this further restriction. But there are other possible relational views which do adopt it and hence which can claim that all classically valid arguments are inference-licensing.

2.4   wishful thinking dispelled

The Frege-Geach problem is a problem in the philosophy of language – the problem of being able to construct a compositional semantics for an expressivist language that leaves the resources to explain the same sorts of things that standard compositional semantic frameworks for non-expressivist languages are able to explain. And it is a problem for any expressivist theory, as applied to metaethics or any other domain. But Cian Dorr [2002] introduced another important problem for metaethical expressivism in particular. In contrast to the Frege-Geach problem, Dorr's *wishful thinking problem* is a problem in epistemology, and in contrast to the Frege-Geach problem, it is a problem for expressivist views in metaethics in particular.[21]

The problem is simple, and again relates to the possibility of 'mixed' normative-descriptive sentences and the possibility of accepting mixed disjunctions while accepting neither disjunct. Suppose that Una believes that either stealing toys from her brother is not wrong or her parents will scold her if they find out, but isn't sure whether it is wrong, and isn't sure whether her parents will scold her if they find out. But after some sincere ethical reflection, she concludes that stealing toys from her brother is in fact wrong. Provided that her ethical reflection has not undermined her justification for her disjunctive belief – and surely there are at least possible cases in which it does not – and provided that her ethical reflection can result in a justified conclusion – and surely it is at least possible that it can – Una would seem to be justified in drawing the conclusion that her parents will scold her for stealing from her brother if they find out. This is just a special case of the observation that except under special circumstances, learning the premises of valid arguments can be a way of gaining justification for accepting their conclusions.

But as Dorr points out, from a standard metaethical expressivist's perspective there is something strange about this case. For when Una comes to accept the conclusion of this reasoning, she is forming an ordinary descriptive belief about the world. But according to a standard metaethical expressivist, when Una comes to accept the key premise of the argument that allows her to perform this inference – her judgment that stealing toys from her brother is wrong – she is simply forming a desire-like attitude. But in general, forming beliefs about how the world is on the basis of your desires is irrational – it is what we call *wishful thinking*. So traditional metaethical expressivists face a dilemma. They must either conclude that wishful

---

[21] For Dorr's original objection, which he described specifically as an objection to *noncognitivism* in order to make clear that it does not apply to just any expressivist view, see Dorr [2002]. For important replies, see Enoch [2004] and Lenman [2004]. For further subsequent discussion, see Schroeder [2010], chapter 9, and Budolfson [2011]. And for the way hybrid expressivism avoids the problem, see Ridge [2007b].

thinking is in fact sometimes rational, or they must place surprising restrictions on the capacity for moral arguments to yield justification for their conclusions.

Dorr is quite explicit, in his original article, that his target includes only views according to which Una's accepting the premise consists in her having a desire-like attitude. So it's clear where the loophole is for tempered theories, and no criticism of Dorr's argument that it doesn't capture them in his net. Tempered theories will get around the problem by holding that insofar as Una comes to accept that stealing toys from her brother is wrong without losing her justified belief that either stealing toys from her brother is not wrong or her parents will scold her if they find out, this is because justifiably coming to accept that stealing toys from her brother is wrong involves justifiably coming to have a new ordinary descriptive belief.

For the hybrid theory, it is simple to cash out this story in more detail.[22] Una's disjunctive normative belief consists, in part, in her having the ordinary descriptive belief that either stealing from her brother does not instantiate K or her parents will scold her if they find out, and in part in her having the desire not to do what instantiates K. Since she already has the desire not to do what instantiates K, therefore, all she lacks in order to come to believe that stealing toys from her brother is wrong, is the ordinary descriptive belief that stealing toys from her brother does not instantiate K. So when she concludes on the basis of her ethical reasoning that stealing toys from her brother is not wrong, what she is really doing is forming this ordinary descriptive belief. But insofar as this reasoning is justified, and insofar as her original disjunctive belief is justified, that gives her justification to conclude that her parents will scold her if they find out – in virtue of how justification works for ordinary descriptive beliefs. No problem about wishful thinking, here.

The relational expressivist avoids the wishful thinking problem for reasons that are similar, though slightly more complicated. Insofar as Una accepts the disjunction, she has some ordinary descriptive disjunctive belief, one of whose disjuncts is that her parents will scold her if they find out that she has stolen toys from her brother, and belief in the other of whose disjuncts would suffice, given her current conditional disapproval state, for her to believe that stealing toys from her brother is not wrong. Call the content of this other disjunct p. Similarly, according to the relational expressivist, provided that Una's conditional disapproval state does not change, in order to come to believe that stealing toys from her brother is wrong, she must come to have an ordinary descriptive belief in the negation of p.

---

[22] See Ridge [2007b] for the hybrid answer to the wishful thinking problem, as realized by his particular hybrid theory, which he calls *ecumenical expressivism*.

So insofar as her conditional disapproval state does not change and both beliefs remain justified, she will be justified in accepting the conclusion of the argument. Moreover, if her conditional disapproval state does change, then she will remain justified in accepting the disjunction only if there is some other ordinary descriptive disjunctive belief that she is justified in having that suffices to play the same role. So regardless, if she comes to justifiably accept that stealing toys from her brother is wrong without losing justification for her disjunctive belief, she will gain justification for believing the conclusion of her argument – and all by means of ordinary justification for ordinary descriptive beliefs, and consequently without any sort of wishful thinking.

In this section we've seen how both forms of tempered expressivism – hybrid expressivism and relational expressivism – avoid problems about wishful thinking. In the last two sections we saw how they deal with core aspects of the Frege-Geach problem. With all of these virtues in common, it is worth comparing these two viws in more detail, and seeing how their commitments compare. I turn to this task, after tracing out some of the antecedents for relational expressivism, in part 3.

3.1   antecedents for relational expressivism

As I've indicated above, I know of no totally clear, explicit, and self-aware endorsements of relational expressivism in the literature.[23] But there are a number of suggestive antecedents for the view, and in this section I'll review a few of the contributions which bear important resemblances to the key ideas of relational expressivism and which have inspired my understanding of it. Then in section 3.2 I'll turn to compare the commitments of relational and hybrid expressivism.

It doesn't take much to make the relational expressivist's semantics look more familiar. Although in principle the goal of the relational expressivist's semantics is to assign each sentence to a relation R between one's ordinary belief state and one's attitudinal state, as we've seen, an important class of such relations can be fully characterized by a function [P] from attitudinal states to propositions, such that your ordinary belief state stands in R to your attitudinal state α just in case it includes an ordinary descriptive belief in the proposition [P]α. So I characterized the relational expressivist's compositional semantics by assigning functions from attitudinal states α to propositions.

But many semanticists either believe that propositions are, or model propositions as, sets of possible worlds. So under this assumption, [P] is a function from attitudinal states to sets of possible worlds. Since the negation of a possible-worlds propositon is its complement, the conjunction of two

---

[23] Though see Toppinen [this volume].

possible-worlds propositions is their intersection, and the disjunction of two possible-worlds propositions is their union, under this assumption we can restate the compositional rules for the relational expressivist as follows: $[\sim P]\alpha = W\backslash[P]\alpha$, $[P\&Q]\alpha = [P]\alpha \cap [Q]\alpha$, and $[P\vee Q]\alpha = [P]\alpha \cup [Q]\alpha$, where W is the set of all possible worlds.

Now it's a familiar fact that a function from some domain to subsets of another domain is interdefinable with a relation between the first domain and elements of the second domain.[24] For example, let $|P| = \{<w,\alpha>: w \in [P]\alpha\}$. Then $[P]\alpha = \{w: <w,\alpha> \in |P|\}$. And it is straightforward to prove, given this definition, that $|P\&Q| = |P| \cap |Q|$ and $|P\vee Q| = |P| \cup |Q|$, and, given the assumption that $[P]\alpha$ is always defined, $|\sim P| = V\backslash|P|$, where V is the set of $<w,\alpha>$ pairs.[25] So given some natural further assumptions, the relational expressivist's semantics is fully definable on the basis of the assignment to each sentence of a set of $<w,\alpha>$ pairs, where these sets compose by the rules of complementation for negation, intersection for conjunction, and union for disjunction.

This should look very familiar. In *Wise Choices, Apt Feelings*, Allan Gibbard assigned each sentence to a set of world-norm pairs, and composed these sets by the operations of complementation for negation, intersection for conjunction, and union for disjunction, as familiar from possible-worlds semantics. So at bottom, the relational expressivist can be thought of as employing a system which, up to the assuption that propositions can be characterized as sets of possible worlds, has essentially the same formal properties as Gibbard's system.

Gibbard, however, is quite clearly not a relational expressivist. This is clear for a couple of reasons. First, it is clear that on Gibbard's own interpretation of his formal system, norms are strictly analogous to worlds, in that rather than corresponding to all possible attitudinal states of mind, they only correspond to *fully decided* attitudinal states, and less than fully opinionated attitudinal states correspond to *sets* of norms between which they are undecided, in the same way that less than fully opinionated belief states correspond to sets of worlds between which they are undecided. (This feature of Gibbard's view becomes even clearer in *Thinking How to Live*, when he introduces talk of *hyperplans*.) But the relational expressivist needs her functions to tell us which belief is necessary and sufficient, given *any* attitudinal state of mind – even less than fully opinionated states.

---

[24] Compare discussion in Dreier [1999] for another application of this same fact, with more detailed discussion.

[25] Proofs: $|P\&Q| = \{<w,\alpha>: w \in [P\&Q]\alpha\} = \{<w,\alpha>: w \in [P]\alpha \cap [Q]\alpha\} = \{<w,\alpha>: w \in [P]\alpha\} \cap \{<w,\alpha>: w \in [Q]\alpha\} = |P| \cap |Q|$. Similarly for $|P\vee Q|$. $V\backslash|P| = \{<w,\alpha>: \sim w \in [P]\alpha\}$. But assuming $[P]\alpha$ is defined for all α, this is just $\{<w,\alpha>: w \in W\backslash[P]\alpha\} = \{<w,\alpha>: w \in [\sim P]\alpha\}\} = |\sim P|$.

Another reason it is clear that Gibbard is not a tempered relational expressivist is that it is clear that he thinks of at least some normative sentences as expressing pure states of norm-acceptance, and hence as being assigned to sets of world-norm pairs that are *world-invariant*, in the sense that every norm that gets paired with any world gets paired with every world. This clearly violates the tempered restriction, and moreover leads to the same problems the tempered relational expressivist sought to avoid. For example, it generates conflations between believing a disjunction and believing one of the disjuncts.[26] So despite superficial similarities, it is clear that Gibbard is not a relational expressivist. Though perhaps some people attracted to Gibbard's formal system should be.

There is also an important relationship between relational expressivism and relativism. To see why, note that a semantics that works by assigning sentences to sets of world-attitudinal-state pairs is equivalent to one that assigns either of some arbitrary pair of values – call them 'T' and 'F' – to sentences relative to worlds and attitudinal states. We simply define $[\![P]\!]^{w,\alpha}$=T iff <w,α>∈|P| and =F otherwise. So this looks like the relativist's idea that a semantics may assign 'truth' values relative to parameters or points of view in addition to worlds.

However, again it is clear that the relational expressivist's interpretation of this system differs from that of the relativist. A proper relativist is going to think that there is some important significance to the fact that we call the values assigned to sentences 'T' and 'F', and correspondingly some important relationship or other between them and our ordinary notions of truth and falsity. But according to the relational expressivist, all it means to say that some sentence 'P' is assigned 'T' relative to w and α is that it is possible for someone who has attitude α to believe that P without her ordinary descriptive belief state having ruled out w as a possibility. This doesn't seem to have anything to do with truth. So despite the formal similarities, I don't think relational expressivism is quite what anyone has had in mind when talking about moral relativism – although it may be a better way of capturing what some people have thought was important.

A distinct but I think also important antecedent for relational expressivism in the literature is James Lenman's [2004] response to Cian Dorr's wishful thinking problem on behalf of the metaethical

---

[26] To see why, suppose that |P| and |Q| are both world-invariant. So each divides the set of attitudinal states into two: those for which you count as having the belief no matter what your ordinary descriptive beliefs, and those for which you don't count, no matter what your ordinary descriptive beliefs. Then |P∨Q|, their union is also world-invariant. It also divides the attitudinal states into two, and in the same way. But the attitudinal states that make you count as believing that P∨Q are just those that come from one of the other two sets. So to believe that P∨Q, you must have an attitudinal state that either suffices for believing that P no matter what your ordinary descriptive beliefs, or suffices for believing that Q no matter what your ordinary descriptive beliefs. This identifies believing that P∨Q with either believing that P or believing that Q.

expressivist. Lenman's article suggests strongly that he believes that his response is one that can be offered by any metaethical expressivist, so it doesn't fit very closely with a relational expressivist interpretation. But his response to the wishful thinking problem has much in common with that I've offered on behalf of the relational expressivist in section 2.4, and thinking about Lenman's article was very fruitful for me in thinking about the prospects for relational expresssivism.

Finally, there is much in common between Michael Ridge's *ecumenical expressivism* ([2006], [2007a], [2007b]) and relational expressivism. Unfortunately for the strict interpretation of Ridge's view as a form of relational expressivism, he has been very explicit in describing it as a hybrid theory, according to which moral sentences express both ordinary descriptive beliefs, and attitudes of approval for certain kinds of idea observers. But it is an important feature of Ridge's view that which attitude is expressed and which belief is expressed both vary freely between speakers, though they are linked as to *co-vary* in an important way – that is, to bear a uniform *relationship* to one another.

In Schroeder [2009] I argued that there are deep and difficult problems with Ridge's account that arise from the fact that he requires both belief and attitude to vary in this way – problems with being able to offer an adequate account of in what sense these states are expressed, problems with attitude ascriptions, and problems with truth ascriptions. But there is a very easy fix for these problems that would preserve all of the core elements of Ridge's positive view that are central to its main attractions. And that would be to adopt relational expressivism. Instead of saying that moral sentences express two states, but that which pair of these states is expressed co-varies between speakers, Ridge could simply say that moral sentences express a uniform but *relational* state of mind, which for different speakers may be satisfied through different ordinary descriptive belief states and correspondingly different approval states. Ridge [2009] begins to move in this direction, and Toppinen [this volume] advocates exactly this move on Ridge's behalf in response to some of my earlier arguments.

In this section I've been going over some of the important antecedents for relational expressivism. With the exception of Toppinen [this volume], I don't believe that any of these authors quite clearly and explicitly adopts a relational expressivist theory, but each bears important resemblances to the relational expressivist perspective, and some of these views – especially Ridge's – might be improved by explicitly embracing it. In the next and final section I'll turn to compare the hybrid and relational ways of tempering expressivism.[27]

---

[27] My own intellectual journal toward thinking about relational expressivism actually started in a different place, by thinking about a different way in which to interpret Gibbard's system of world-norm pairs and trying to interpret the views described by Field [2000], [2009].

3.2   comparing

As we saw in part 2 of this paper, hybrid and relational expressivism have a great deal in common; indeed, hybrid expressivism can be thought of as a special case of relational expressivism, under its most general characterization, and the familiar compositional methods of the hybrid expressivist can be thought of as a special case of a more general method which applies to other forms of relational expressivism. But they also have importantly different commitments.

The distinctive commitments of the hybrid expressivist arise because she holds that there is a *particular* ordinary descriptive belief state that you need to be in, to accept some particular sentence, and a *particular* desire-like attitude that you must have, in order to accept any sentence involving a normative term. It's a consequence of the first of these commitments that hybrid expressivism amounts to a kind of metaethical *realism*. If what motivated you to be an expressivist in the first place was the prospect of getting to avoid having to say what it is to be wrong, this sort of view therefore doesn't deliver. It is most attractive, in contrast, for realists who hope to be able to co-opt some of the attractions of expressivism in accounting for moral motivation, or perhaps in dealing with Open Question phenomena.

A slight complication arises for hybrid expressivists who claim that the ordinary descriptive belief state that you need to be in, in order to accept a particular normative sentence, varies from context to context as a result of the contribution of the normative term. Depending on their other features, contextualist views can end up looking substantially irrealist, if they end up allowing for enough contextual variation that it is hard to maintain that there is any single subject matter of normative discourse. But the same points go for such views as before; they must still be able to say, for any particular context, what you need to believe, in order to accept a normative sentence in that context.

The other important commitment of hybrid theories is that there is some particular desire-like attitude that you must have, in order to accept any sentence involving a particular normative term. It's a consequence of this commitment that hybrid theorists must allow that there is some desire-like attitude that has been shared by everyone who has ever had any view about whether anything is wrong or not – no matter what view they had. This is a substantial empirical hypothesis, and if it were true, that would seem to explain much of the motivational import of normative claims even if the hybrid theory were false.

Again, for hybrid theories according to which the desire-like attitude that you need to be in can vary from context to context, things are slightly more complicated. But again, any such view is going to be committed to an empirical hypothesis to the effect that anyone who has ever had a view about what is wrong has had whatever desire-like attitude was required in her context, in order to do so. But once we

24

accept this empirical hypothesis, there will be an ordinary, non-expressivist, contextualist view which assigns normative sentences to the same descriptive contents as the hybrid theory does but does not build in any hybrid expressivist component, and there is little that the hybrid theory adds to the explanation of moral motivation, over and above this empirical hypothesis. So hybrid theories have commitments which give them only subtle advantages, if any, over counterpart views which assign only descriptive contents to normative sentences.[28]

In contrast, relational expressivism comes with very different commitments that align it much more closely with the traditional irrealist ambitions of familiar views in the noncognitivist tradition. For the relational expressivist, there is no particular ordinary descriptive belief or desire-like attitude that you need to have, in order to accept a normative sentence; you must simply instantiate the right relationship between these two aspects of your overall state of mind. This absolves the relational expressivist of needing to be able to say what property moral terms pick out, and it allows for people with very different attitudes to all count as having views about what is wrong or not.

The strong commitments of the relational expressivist, in contrast, come in her ideas that there are no normative sentences that express purely attitudinal states of mind, and her commitment that the relation her account appeals to must be belief-monotonic.

Consideration of simple normative terms like 'good', 'wrong', and 'ought' can lend plausibility to the idea that normative terms in natural language don't purely constrain one's attitudinal state, because the applicability of these terms in general depends not only on the norms, but also on the facts. Money may be good if it can buy happiness but not if it only creates stress, for example, and not giving Jack money may be wrong if you've promised to pay him, but not if you haven't. This makes it seem plausible that natural-language normative terms really only constrain the relationship between norms and beliefs, rather than norms alone.

But I think this intuitive idea falls short of motivating the thesis the relational expressivist needs. What she needs is not just the thesis that words like 'good', 'wrong', and 'ought' only constrain combinations of belief with attitude, but that no term *could* constrain only attitude. But moral philosophers at least since Aristotle have deliberately tried to introduce terms like 'intrinsically good' and 'intrinsically wrong' precisely in order to make claims that are fully independent of non-moral contingencies like those mentioned in the last paragraph.

---

[28] For further discussion of this point, see Schroeder [2009], especially sections 8 and 9.

The restriction to belief-monotonic relations is also significant. As I noted before, my illustrative example of thinking of categorical plans as a relationship between contingency plans and prosaic belief is misleading, because this relation is belief-*non*monotonic. To make good on her view, a relational expressivist must reject this feature of the example and articulate a view on which the relation is belief-monotonic. Ridge's ecumenical expressivism, reinterpreted as relational expressivism, would provide just such a picture. Perhaps there are others.[29]

3.3   commitments in the philosophy of mind

As I've shown in this paper, there are a number of important advantages which could accrue to the tempered relational expressivist, if she is able to discharge these challenges. This makes relational expressivism worth taking seriously and exploring further, in keeping with my second announced goal for this paper. And it also serves my first announced goal in an important way. Typical expressivist views in metaethics start with the idea that there are two very different aspects of our psychology, each of which can be expressed by declarative sentences, and then need to postulate third or fourth aspects of our psychology expressible by declarative sentences, in order to have non-conflating psychological states to be expressed by complex sentences — in particular, by mixed disjunctions.

The question of how far an expressivist view can get without postulating third or fourth distinct psychological attitudes can be framed as the question of how far we can get by supposing that every state expressible by a declarative sentence can be fully characterized as a relation between two aspects of one's psychology. So thinking of the states expressed by declarative sentences as themselves relations between distinct psychological states is the most general way of exploring how far we can get without adopting strong commitments in the philosophy of mind by going in for more psychological attitudes, just in order to account for complex thoughts. The possibility of tempered relational expressivism shows that we can get relatively far, provided that our commitments in the philosophy of mind come up front — in the form of tempering — provided that we are able to adopt the right sort of view of the relational state.[30]

---

[29] One reason this constraint is so important is that I think it precludes applying the strategy of tempered relational expressivism to expressivism about epistemic expressions like 'might', 'probably', the indicative conditional, and even the deliberative 'ought'. Because of their distinctive epistemic component (see Ross and Schroeder [unpublished] for elaboration), I don't believe that any of these terms will be capturable with belief-monotonic relations.

[30] Special thanks to Ben Lennertz, Hartry Field, Jamie Dreier, Max Barkhausen, Daniel Boisvert, Teemu Toppinen, Mark van Roojen, Ryan Millsap, Michael Ridge, and Russ Shafer-Landau, and to audiences at the Wisconsin Metaethics Workshop, Duke University, and Stanford University.

references

Ayer, A.J. [1936]. *Language, Truth, and Logic*. New York: Dover.

Barker, Stephen [2000]. 'Is Value Content a Component of Conventional Implicature?' *Analysis* 60(3): 268-279.

Blackburn, Simon [1984]. *Spreading the Word*. Oxford: Oxford University Press.

_____ [1993]. *Essays in Quasi-Realism*. Oxford: Oxford University Press.

_____ [1998]. *Ruling Passions*. Oxford: Oxford University Press.

Boisvert, Daniel [2007]. 'Expressive-Assertivism.' Forthcoming in *Pacific Philosophical Quarterly*.

Budolfson, Mark [2011]. 'Noncognitivism and Rational Inference.' *Philosophical Studies* 153(2): 243-259.

Copp, David [2001]. 'Realist-Expressivism: A Neglected Option for Moral Realism.' *Social Philosophy and Policy* 18: 1-43.

_____ [2009]. 'Realist-Expressivism and Conversational Implicature.' In *Oxford Studies in Metaethics*, Volume 4, ed. Russ Shafer-Landau. Oxford: Oxford University Press, 167-202.

Dorr, Cian [2002]. 'Non-cognitivism and Wishful Thinking.' *Noûs* 36(1): 97-103.

Dreier, James [1999]. 'Transforming Expressivism.' *Noûs* ….

_____ [2006]. 'Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution.' *Oxford Studies in Metaethics*, volume 1. Oxford: Oxford University Press, 217-233.

Enoch, David [2003]. 'How Noncognitivists Can Avoid Wishful Thinking.' *Southern Journal of Philosophy* 41: 527-545.

Field, Hartry [2000]. 'A Prioricity as an Evaluative Notion.' in P. Boghossian and C. Peacocke, eds., *New Essays on the A Priori*. Oxford: Oxford University Press, 117–49.

_____ [2009]. 'Epistemology Without Metaphysics.' *Philosophical Studies* 143(2): 249–90.

Geach, P.T. [1965]. 'Assertion.' *Philosophical Review* 74: 449-465.

Gibbard, Allan [1990]. *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.

_____ [2003]. *Thinking How to Live*. Cambridge: Havard University Press.

Hay, Ryan [2011]. 'Hybrid Expressivism and the Analogy Between Pejoratives and Moral Language.' *European Journal of Philosophy*, publish Online First, February 23, 2011.

Horgan, Terry, and Mark Timmons [2006]. 'Cognitivist Expressivism.' In Horgan and Timmons, eds., *Metaethics after Moore*. Ridge, Michael [2006]. 'Ecumenical Expressivism: Finessing Frege.' *Ethics* 116(2): 302-336.

Lenman, James [2003]. 'Noncognitivism and Wishfulness.' *Ethical Theory and Moral Practice* 6: 265-274.

Potts, Christopher [2005]. *The Logic of Conventional Implicature*. Oxford: Oxford University Press.

Ridge, Michael [2006]. 'Ecumenical Expressivism: Finessing Frege.' *Ethics* 116(2): 302-336.

_____ [2007a]. 'Ecumenical Expressivism: The Best of Both Worlds?' *Oxford Studies in Metaethics* 2: 51-76.

_____ [2007b]. 'Epistemology for Ecumenical Expressivists.' *Proceedings of the Aristotelian Society* suppl. volume 81: 83-108.

_____ [2009]. 'Moral Assertion for Expressivists.' *Philosophical Issues* 19: 182-204.

Ross, Jacob, and Mark Schroeder [unpublished]. 'Reversibility or Disagreement.' Unpublished manuscript.

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

_____ [2009]. 'Hybrid Expressivism: Virtues and Vices.' *Ethics* 119(2): 257-309.

_____ [2010]. *Noncognitivism in Ethics.* New York: Routledge.

_____ [forthcoming]. 'Hard Cases for Combining Expressivism and Deflationist Truth: Conditionals and Epistemic Modals.' Forthcoming in a volume of essays on pragmatism and deflationism, edited by Michael Williams and Steven Gross.

_____ [unpublished]. 'Attitudes and Epistemics.' Unpublished paper.

Searle, John [1962]. 'Meaning and Speech Acts.' *Philosophical Review* 71:423-432.

Toppinen, Teemu [this volume]. 'Belief in Expressivism.' Forthcoming in Russ Shafer-Landau, ed., *Oxford Studies in Metaethics*, volume 8.

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

van Roojen, Mark [1996]. 'Expressivism and Irrationality.' *The Philosophical Review* 105(3): 311-335.

Yalcin, Seth [2007]. 'Epistemic Modals.' *Mind* 116(4): 983-1026.