

Advice for Noncognitivists*

Malte Willer

Abstract

Metaethical noncognitivists have trouble arriving at a respectable semantic theory for moral language. The goal of this paper is to make substantial progress toward demonstrating that these problems may be overcome. Replacing the predominant expressivist semantic agenda in metaethics with a dynamic perspective on meaning and communication allows noncognitivists to provide a satisfying analysis of negation and other constructions that have been argued to be problematic for metaethical noncognitivism, including disjunctions. The resulting proposal preserves some of the key insights from recent work on the semantics of expressivism while highlighting the widely neglected early noncognitivists' sympathies to the kind of dynamic story I intend to tell here. A comparison between the advertised dynamic semantic story and current proposals that treat expressivism as a pragmatic rather than semantic theory about moral language is provided.

1 The Plot

It is a well-worn story that despite all their differences, modern-day metaethical expressivists (most notably Blackburn (1984, 1988) and Gibbard (1990, 2003)) and their early noncognitivist cousins (Ayer (1936), Hare (1952), and Stevenson (1944, 1963)) struggle alike in arriving at a theoretically satisfying semantics for moral and descriptive language (the Frege-Geach problem). The goal of this paper is to demonstrate that synthesizing insights from both traditions in a dynamic semantics for natural language delivers a noncognitivist perspective on the language of morals that avoids the semantic problems of its predecessors.

Contemporary noncognitivism in metaethics stresses the role of language as a means for *expressing* states of mind. The difference between moral sentences such as 'Stealing is wrong' and descriptive ones such as 'Grass is green,' so the story goes, is to be located in the *kind* of state they are designed to express: to say that grass is green is to express a state of belief; to say that stealing is wrong, in contrast, is to express a certain conative, desire-like state. From a naive perspective, however, language is first and foremost a tool for *coordinating* rational activity and this, one would think, highlights its role as a means for coordinating—rather than simply expressing—states of mind. While noncognitivist theories of all stripes recognize the coordinative dimension of the language of morals—even Ayer's account with its heavy emotivist bent, for instance, stresses that moral terms are designed to arouse feelings and so to stimulate action—it plays no role in their *semantic*

*Published in *Pacific Philosophical Quarterly*, 98(S1): 174–207, 2017

story about the language of morals. My claim here is that taking coordination rather than expression as a starting point for semantic theorizing allows us to make some real progress toward a satisfying noncognitivist semantics for natural language.

Many linguists and philosophers are inclined to follow [Stalnaker \(1978\)](#) in treating coordination as a pragmatic rather than semantic phenomenon, and I should stress that highlighting the coordinative dimension of language is *not* meant to be a direct empirical or conceptual argument against this tradition. It does, however, suggest that instead of following the common expressivist path by modeling the semantic value of a sentence in terms of the attitude it expresses (or some representation thereof), one may take a dynamic path and model the semantic value of a sentence in terms of its role in coordinating states of mind, that is, in terms of the *changes* that the sentence is designed to induce in the recipient’s attitudes. Semantic values thus become operations on attitudes, and we can say all this while preserving the key difference between descriptive and moral language that metaethical noncognitivists care so much about: descriptive sentences, we may say, have *cognitive* content in the sense that they are designed to induce ordinary beliefs; moral sentences, in contrast, have *noncognitive* content in the sense that they are designed to induce desire-like states.¹

The proposed shift from an expressivist to a dynamic semantics, or so I shall argue here, does more than putting a noncognitivist gloss on moral language understood as a medium for coordinating states of mind. It is a well-worn story that any successful noncognitivist semantics for some target language must not only be compositional—yield a method for computing the semantic value of any well-formed construction on the basis of the semantic values of its components—but also deliver a satisfying conception of key semantic notions such as the one of inconsistency between sentences. The central innovation of the dynamic framework developed here is that it not only takes the first hurdle with grace but also revives an attractive strategy for handling the second challenge that is explored, but ultimately dismissed as unworkable in an expressivist setting, by [Schroeder \(2008a, 2008b\)](#).² In line with the mainstream expressivist tradition, the upcoming story will explain semantic relations between sentences in terms of relations between attitudes—two sentences ϕ and ψ , for instance, are inconsistent just in case they are designed to induce inconsistent (cognitive or noncognitive) attitudes. In contrast with the expressivist tradition, however, it is in a position to explain inconsistency between noncognitive states in the same way we would ordinarily explain why two beliefs are inconsistent: as instances of the same attitude toward (classically) inconsistent contents. While the prospects of implementing this explanatory strategy in an expressivist setting are bleak—as Schroeder observes—it can be pursued without negative side effects once logical connectives are modeled dynamically as operators on update functions rather than—as expressivism would have it—on states of mind.

¹The distinction between cognitive and noncognitive content is not meant not be exhaustive. Certain sentences arguably have *inquisitive* content in the sense that they raise issues while others have *attentive* content in virtue of their potential to draw attention to certain possibilities (as discussed by [Ciardelli et al. \(2009\)](#), [Roelofsen \(2013\)](#), and [Willer \(2015\)](#), among others).

²It is in virtue of this important innovation that the framework developed here differs in spirit and at crucial moments of detail from the dynamic approaches to the language of norms by [Alwood \(forthcoming\)](#), [Charlow \(2015\)](#), and [Starr \(forthcoming\)](#).

My strategy for explaining why all of this makes sense and is more than a merely technical result is as follows. §2 outlines the scope as well as the limits of mainstream metaethical expressivism. §3 motivates a dynamic perspective on meaning and communication and explains how it captures the role of moral language as a medium for coordinating desire-like states that is emphasized by early noncognitivist authors such as Ayer and Stevenson. The resulting proposal preserves some of the key insights from recent work on expressivism but offers a formally rigorous solution to its main shortcoming—most notably, it demonstrates how noncognitivists can explain logical inconsistency in terms of what [Schroeder \(2008a, 2008b\)](#) labels ‘*A*-type inconsistency’ between states of mind without negative side effects—and thus makes real progress toward a compositional semantics for natural language that is sympathetic to the noncognitivist tradition in metaethics (§4). The final §5 compares the advertised solution to the Frege-Geach problem for noncognitivism with recent trends in the expressivist literature, including the suggestion to combine a classical truth-conditional semantics with an expressivist pragmatics.

2 Expressivism and Its Limits

This is not the place to offer a comprehensive history of the Frege-Geach problem for metaethical noncognitivism.³ A better starting point is to look at how the trouble unfolds for the most prominent expressivist incarnation of the noncognitivist agenda, and to see what shape a successful solution could take. The hallmark of an expressivist semantics for some language \mathcal{L} , as I will understand it here, is that it assigns semantic values to sentences of \mathcal{L} in terms of (abstract representations of) the states of mind those sentences express. The descriptive sentence ‘Grass is green,’ for instance, expresses a belief: a representational state that is evaluable for truth or falsity. The moral sentence ‘Stealing is wrong,’ in contrast, expresses some moral attitude: one that is desire-like and has the capacity to motivate action. Logical connectives are interpreted as functions from states of mind to states of mind: the attitude expressed by, say, ‘Stealing is not wrong’ is determined by the attitude expressed by ‘Stealing is wrong’ together with a general semantic rule for how the negation operator maps an input to an output state. And so the logical relations between elements of \mathcal{L} , most notably the ones of entailment and inconsistency, are to be explained in terms of relations between the states of mind expressed by those sentences.⁴ The Frege-Geach problem for this framework, should there turn out to be any such problem at all, is that expressivists cannot offer a theoretically satisfying explanation of the semantic relations between elements of \mathcal{L} in terms of the states of mind they are supposed to express. Let me explain.

Start with the observation that (1) and (2) are inconsistent:

- (1) Stealing is wrong.
- (2) Stealing is not wrong.

³And it is done elsewhere by [Schroeder \(2008c, 2010\)](#). The classical discussions are by [Geach \(1960, 1965\)](#) and [Searle \(1962\)](#). Geach credits [Frege \(1919\)](#) for the original insight.

⁴At a minimum, this is the standard take on the semantic agenda of expressivism: see the discussions by, e.g., [Dreier \(2009\)](#), [Horgan and Timmons \(2006\)](#), [Rosen \(1998\)](#), [Schroeder \(2008a, 2008b\)](#), [Wedgwood \(2007\)](#), and [Unwin \(2001\)](#). [Charlow \(2014\)](#) and [Silk \(2013\)](#) develop non-standard outlooks on the semantics of expressivism, which I will discuss in more detail in the concluding section of this paper.

Expressivist semantics has it that (1) and (2) are inconsistent, not in virtue of their incompatible truth-conditions (the classical explanation), but rather in virtue of the relation between the attitudes these sentences express. The specific proposal, of course, is that (1) and (2) are inconsistent because the attitudes they express stand in the distinguished relation of inconsistency—one that we are well-advised to distinguish from other clashes of attitude such as the tension between believing that grass is green and believing that one does not believe that grass is green.⁵ The resulting challenge for an expressivist analysis of negation then is that it must explain why the attitudes expressed by (1) and (2) are inconsistent rather than simply stipulate that they are. And this is something that the classical expressivist frameworks from Blackburn and Gibbard simply do not deliver—or so Dreier (2006, 2009), Schroeder (2008a, 2008b, 2008c), Schueler (1988), Sinnott-Armstrong (2000), and Unwin (1999, 2001) have all argued.

Schroeder (2010) helpfully distinguishes the problem of arriving at a satisfying explanation of the inconsistency between (1) and (2) from the one of arriving at a compositional expressivist semantics for natural language (see also Charlow 2014). Compositionality requires a method for computing the semantic value of any well-formed construction of \mathcal{L} on the basis of the semantic values of its components. An expressivist semantics for some language \mathcal{L} , so the observation goes, may be compositional in this sense but still fail to adequately explain the logical relations between its elements. The proposal by Gibbard (1990, 2003), for instance, models states of mind using pairs consisting of a possible world and some ‘hyperplan’—a plan that either forbids or allows a certain kind of action. The semantic value of (1), $\llbracket(1)\rrbracket$, is the set of pairs $\langle w, \pi \rangle$ such that ‘according to π , stealing is forbidden’ is true at w . Negation, conjunction, and so on receive their classical Boolean interpretation, and so the semantic value of (2), $\llbracket(2)\rrbracket$, is the set of pairs $\langle w, \pi \rangle$ such that ‘according to π , stealing is allowed’ is true at w . The semantics is obviously compositional—all we have done is to add some complexity to the notion of a possible world that is familiar from classical semantics—but by itself does not meet the challenge of explanatory adequacy: there remains the question *why* the states of mind modeled by $\llbracket(1)\rrbracket$ and $\llbracket(2)\rrbracket$ are inconsistent. For while it is perfectly plausible to suggest that (1) expresses the attitude of disapproval of stealing while (2) expresses the attitude of tolerance of stealing, it amounts to nothing more than saying that (1) and (2) express distinct attitudes toward the same kind of action that—as far as the analysis so far is concerned—might very well be logically unrelated.

Gibbard (2003) explains negation in terms of *disagreement*: to accept a sentence of the form ‘ $\neg\phi$ ’—to be in the state of mind expressed by that sentence—is to disagree with the attitude expressed by ϕ . Insofar as negation thus understood is maintained to adequately characterize the states of mind expressed by negated sentences such as (2), however, the relevant notion of disagreement is a technical one in need of further explanation. To wit, someone who is thoroughly agnostic or undecided about the moral status of stealing in a sense disagrees with the attitude of disapproval of stealing but does not thereby accept (2). The notion of disagreement that figures in an expressivist analysis of negation must thus go beyond the one of just *failing* to have the attitude expressed by

⁵As van Roojen (1996) observes, not just any clash of attitude will serve as a foundation for semantic inconsistency. Intuitively, ‘Grass is green’ expresses the belief that grass is green while ‘I do not believe that grass is green’ expresses the belief that one does not believe that grass is green. There is a sense in which these attitudes clash, but the sentences used to express them are not semantically inconsistent (though an utterance of ‘Grass is green and I do not believe that grass is green’ is pragmatically marked).

what is negated (see Dreier (2006) and Unwin (1999, 2001)). The worry then is that if disagreement can only be explained by appealing to two basic attitudes—tolerance and disapproval, for instance—the inconsistency between disagreeing states of mind remains a pure stipulation.

Gibbard (2013) insists that expressivists have every right to base their semantics for the language of morals on a primitive notion of inconsistency between basic attitudes (see also Baker and Woods 2015) but here I am interested in the prospects of an explanatory strategy pursued, but ultimately deemed to be unsuccessful in an expressivist setting, by Schroeder (2008a, 2008b). Start with the classical explanation of the inconsistency between the belief that grass is green and the belief that grass is not green. The sentences ‘Grass is green’ and ‘Grass is not green’ are inconsistent since they have incompatible truth-conditions. But two beliefs are inconsistent whenever their contents are inconsistent—belief, as we may say, is ‘inconsistency-transmitting’ in the following sense:

Inconsistency-Transmitting Attitudes An attitude *A* is *inconsistency-transmitting* just in case two instances of *A* are inconsistent just in case their contents are inconsistent.

Not all attitudes are inconsistency-transmitting: there is, for instance, nothing inconsistent about wondering whether grass is green and wondering whether grass is not green. But belief certainly is, and so it is not surprising at all that the belief that grass is green is inconsistent with the belief that grass is not green.

The important observation is that the classical story about belief is one that expressivists may happily accept. In particular, they need not take any issue with assigning to descriptive sentences truth-conditions and thus with the classical account of inconsistency between descriptive contents. The suggestion then is that expressivists generalize the classical explanation of disagreement between belief states so that it also covers the inconsistency between the attitudes expressed by some moral sentence and its negation.⁶ Of course, they cannot simply say that moral sentences express beliefs. But under the reasonable assumption that conative, desire-like states may be inconsistency-transmitting as well, it is open to expressivists to say that just as ‘Grass is green’ and ‘Grass is not green’ are inconsistent because they express a single *cognitive* attitude toward inconsistent contents, so are ‘Stealing is wrong’ and ‘Stealing is not wrong’ because they express a single *noncognitive* attitude toward inconsistent contents. For ease of reference, I will follow here Schroeder and say that we have an ‘*A*-type inconsistency’ between two attitudes whenever they are instances of an inconsistency-transmitting attitude toward inconsistent contents.

Implementing the explanatory strategy I have outlined here—accounting for the inconsistency between the states of mind expressed by (1) and (2) in terms of *A*-type inconsistency—is not completely trivial even if we assume (as I will do here) that moral sentences express inconsistency-transmitting attitudes. Above all, our expressivist semantics for negation must actually *predict* that (1) and (2) express instances of the same

⁶An alternative strategy is pursued by Silk (2015) and Starr (forthcoming), who (while parting ways at crucial moments of detail) explain disagreement between conative states in terms of preferential incoherence rather than in terms of inconsistency of their contents. Part of the exercise pursued here is to demonstrate that this move—irrespective of its virtues or vices—is strictly optional for those with noncognitivist sympathies.

attitude toward inconsistent contents, and their contents must be descriptive so that we can appeal to the well-established classical account of inconsistency between contents. To illustrate: the interpretation considered earlier on which (1) expresses the attitude of disapproval of stealing while (2) expresses *tolerance* of stealing will not deliver the required result since all we have here are two distinct attitudes toward a single content (in Schroeder’s terminology, any inconsistency between these attitudes can only be of ‘type *B*’ and a pure matter of stipulation). Nor will the proposal to interpret (1) as expressing a state of accepting a plan for action that *forbids* stealing while (2) expresses a state of accepting a plan for action that *permits* stealing: here we have two instances of a single attitude, but the alleged inconsistency of their contents—a plan that forbids stealing and a plan that permits stealing—is no less in need of an explanation than the one between the attitudes of disapproval and tolerance of stealing. These obstacles notwithstanding, we now have identified a constraint on an expressivist semantics for negation that, together with some minimal assumptions about the kind of state that moral sentences express, promises a satisfying account of why (1) and (2) express inconsistent attitudes and thus, *mutatis mutandis*, of why the sentences themselves are inconsistent.

In fact, it does not take much to come up with a proposal that will do the trick. As Schroeder observes, we may continue to think of logical connectives as functions from attitudes to attitudes but let them play their more familiar role as operators on descriptive contents. Moral sentences express some inconsistency-transmitting pro-attitude, say the one of *being for*, the moral predicate contributing to that state’s descriptive content. The sentence ‘Stealing is wrong,’ for instance, expresses a state of being for blaming for stealing. If ϕ expresses some pro-attitude FOR toward some descriptive content α , then the negation of ϕ expresses FOR($\neg\alpha$). ‘Stealing is not wrong,’ for instance, expresses the state of being for not blaming for stealing. Similarly, if ϕ expresses FOR(α) and ψ expresses FOR(β), then the conjunction of ϕ and ψ expresses FOR($\alpha \wedge \beta$). And so on. The immediate advantage of the strategy: the logical inconsistency between moral sentences is guaranteed to reduce to the familiar inconsistency between the descriptive contents of the noncognitive states those sentences express. And that is just all we need to explain why (1) and (2) are inconsistent: assuming that being for is an inconsistency-transmitting attitude, and since the semantics for negation predicts that (1) and (2) express this attitude toward inconsistent (descriptive) contents, it is not surprising that the attitudes expressed by these sentences disagree with each other—any more than the inconsistency between the belief that grass is green and the belief that grass is not green.

Nothing about the strategy outlined here requires giving up on Gibbard’s proposal to analyze negation as disagreement, and we can also continue to think of the disagreeing attitudes expressed by (1) and (2) as tolerance and disapproval of stealing. Tolerance and disapproval of stealing, it turns out, are instances of the same inconsistency-transmitting attitude—the one of *being for*—toward inconsistent descriptive contents—*blaming for stealing* and *not blaming for stealing*. At this level of analysis, then, negation maps an attitude to a disagreeing attitude by negating its descriptive content. Our analysis of the disagreement relation no longer appeals to two distinct attitudes that may very well be logically unrelated but turns out to be a case of *A*-type inconsistency—which just seems to be the right kind of relation to ground the inconsistency between a sentence and its negation in an expressivist setting.

Unfortunately, the strategy just outlined, while initially attractive, runs into substantial trouble. Schroeder in fact presents several sophisticated objections (more on them below) but the basic problem is easy enough to identify. Simply observe that (2) is not only inconsistent with (1) but also with (3), as is (4):

- (3) Stealing is wrong and grass is green.
- (4) Grass is not green.

If our hope is to explain the inconsistency between these sentences in terms of *A*-type inconsistency between attitudes, then (2) and (3) must express the same attitude toward inconsistent contents, and for the same reason (3) and (4) must express the same attitude toward inconsistent contents. It follows that ‘Stealing is wrong’ expresses the same kind of attitude as does ‘Grass is green’ (though toward different contents), and in general there will be no difference in the kind of attitude that moral and descriptive sentences express. Assuming again that descriptive sentences express beliefs, and since expressivists cannot accept the conclusion that moral sentences express beliefs, that is just to say that belief turns out to be a state of being for—a conclusion that noncognitivists, who insist on a difference in kind between cognitive and noncognitive states, are unlikely to be able to accept either. So much for the bad news for noncognitivists.⁷

The good news is that there is more than one way to implement the strategy of accounting for semantic inconsistency in terms of *A*-type inconsistency. Motivating and elaborating this implementation in detail is a job better reserved for the next section, but let me briefly indicate where we went wrong before and how we can do better. To get the diagnosis on the way, consider the semantic value that the proposal by Gibbard (1990, 2003) assigns to (3): ‘Stealing is wrong’ denotes again the set of world-norm pairs $\langle w, \pi \rangle$ such that ‘according to π , stealing is forbidden’ is true at w while ‘Grass is green’ denotes the set of world-norm pairs $\langle w, \pi \rangle$ such that grass is green at w , and so $\llbracket(3)\rrbracket$ is just the intersection of the two sets. Gibbard’s semantics can thus be reasonably understood as predicting that (3) expresses a *conjunction* of attitudes, one being noncognitive while the other being cognitive in kind, rather than a single attitude toward some conjunctive content. And this is not only intuitive but—at least on first sight—also opens up a path toward an *A*-type explanation of the inconsistency of (3) with (1) and (4) that preserves the difference between cognitive and noncognitive states: to be in the state of mind expressed by (3) is to have a noncognitive attitude that is *A*-type inconsistent with the one expressed by (1)—hence (1) and (3) are inconsistent—and to have a cognitive attitude that is *A*-type inconsistent with the one expressed by (4)—hence (3) and (4) are inconsistent.

But attractive as it is, saying that (3) expresses a conjunction of attitudes—rather than a single attitude toward some conjunctive content—clashes with the key manoeuvre that first allowed us to provide a Schroeder-style *A*-type explanation of why the attitudes expressed by (1) and (2) are inconsistent: *that* strategy crucially relies on the idea that

⁷Note here that the basic problem does not depend on the details of one’s favorite expressivist *A*-type solution to the negation problem: it affects *any* proposal holding ϕ and ψ to be inconsistent just in case they express instances of the same single attitude toward inconsistent contents. Schwartz and Hom (2015), for instance, respond to the negation problem by adding quantificational structure to the content of noncognitive attitudes. This approach differs from the one pursued by Schroeder (2008a, 2008b) at crucial moments of detail, but both accounts are equally ill-equipped to explain inconsistency *A*-type style once we consider combinations of moral and descriptive sentences.

connectives are operators on the *contents* of the attitudes they take as their argument. For consider again a (generalized version of) Schroeder’s crucial definition of negation—that whenever ϕ expresses some attitude ATT toward some descriptive content α , then the negation of ϕ expresses ATT($-\alpha$). Then clearly our definition of negation does not apply in case ϕ expresses some combination of attitudes toward distinct contents—it requires that its complement expresses a single attitude toward some (though perhaps quite complex) content—and thus supposing that (3) expresses a conjunction of attitudes toward distinct contents leads to a Frege-Geach problem: we are at loss when it comes to deriving the semantic values of the negation of (3) in a compositional fashion. The dilemma for expressivists, in brief, then is that the attractive analysis of connectives as operating on the contents of attitudes clashes with what seems to be required to uphold the distinction in kind between cognitive and noncognitive states: that combinations of such states, rather than a single state toward a certain content, are expressed by combinations of moral and descriptive sentences.

But we can do better by replacing the question of what attitude a sentence expresses with the one of what it would take for an agent to come to accept that sentence. Coming to accept a conjunction such as ‘Stealing is wrong and grass is green’ is to adopt a belief *and* some moral attitude—the moral attitude one comes to adopt if one comes to accept ‘Stealing is wrong’ *and* the belief that one comes to adopt if one comes to accept ‘Grass is green.’ This change in perspective may not sound like much but the difference is in fact substantial: we can now say that semantic values are no longer, as expressivists would have it, modeled as (abstract representations) of states of mind but rather as *dynamic operations* on such states of mind. Logical connectives, accordingly, now modify dynamic state change operations—conjunction, for instance, simply amounts to function composition—and it turns out that such a change of perspective allows us to offer a uniform perspective on negation with some real explanatory purchase.

Coming to accept a negation such as ‘Stealing is not wrong’ or ‘It is not the case that stealing is wrong and that grass is green,’ we shall say, is to adopt a state of mind that *rules out* coming to accept what is negated. This proposal rings familiar from Gibbard’s analysis of negation as disagreement, and so we better make sure that our notion of ruling out is not just another primitive relation between basic attitudes. And indeed, the notion of ruling out coming to accept ϕ is *not* a primitive one but amenable to an *A*-type explanation: to rule out coming to accept ϕ is to be in a state in which one could not come to accept ϕ without adopting an inconsistent state of mind. Moreover—and this is the important wrinkle—the relevant notion of inconsistency is an *A*-type inconsistency: a state of mind is inconsistent if, and only if, it involves two instances of a single inconsistency-transmitting attitude toward inconsistent (descriptive) contents. Even without demonstrating that the details of this proposal work out—and I will—there is reason for optimism that a dynamic analysis of negation can do the required explanatory lifting for noncognitivists without the unwanted side-effects that expressivism brings in its wake.

Whatever the technical merits of this proposal, an initial worry is that shifting toward a dynamic perspective on meaning and communication will be just that: a technical fix. In what follows, I hope to show that dynamic semantics in fact offers a very intuitive perspective on moral language, but it is also worth pointing out that the general momentum of the proposal—the idea of moral language being a tool for *coordinating*, rather than simply expressing, moral attitudes—is very prominent in the early noncognitivists’ writ-

ings. [Stevenson \(1944\)](#), to give the perhaps most obvious example, suggests an analysis of ‘This is good’ as a conjunction of a self-report of approval with an imperative to adopt the same approving attitude. How to analyze imperatives semantically is, of course, a very thorny issue, and on some views such constructions have representational content after all (see the discussion by [Portner \(forthcoming\)](#) for an overview of the current state of the art). But Stevenson has a very clear view on what his proposed analysis of ‘This is good’ entails:

Its component imperative, being never used *merely* as an introspective report, renders unambiguously explicit the fact that ‘good’ is used not only in expressing beliefs about attitudes, but in strengthening, altering, and guiding the attitudes themselves. (pp. 23 f.)

What matters for Stevenson, then, is that moral terms are not merely used for expressing beliefs (or moral attitudes for that matter) but are designed to affect the conative, desire-like attitudes of the discourse participants.

[Ayer \(1936\)](#) is commonly labeled a pure emotivist, and it does not take much to see why:

The presence of an ethical symbol in a proposition adds nothing to its factual content. Thus if I say to someone, ‘You acted wrongly in stealing that money,’ I am not stating anything more than if I had simply said, ‘You stole that money.’ In adding that this action is wrong I am not making any further statement about it. I am simply evincing my moral disapproval of it. It is as if I had said, ‘You stole that money,’ in a peculiar tone of horror, or written it with the addition of some special exclamation marks. The tone, or the exclamation marks, adds nothing to the literal meaning of the sentence. It merely serves to show that the expression of it is attended by certain feelings in the speaker.

If now I generalise my previous statement and say, ‘Stealing money is wrong,’ I produce a sentence which has no factual meaning—that is, expresses no proposition which can be true or false. It is as if I had written ‘Stealing money!!’—where the shape and thickness of the exclamation marks show, by a suitable convention, that a special sort of moral disapproval is the feeling which is being expressed. It is clear that there is nothing said here which can be true or false. (p. 110)

However, Ayer is clearly sympathetic to the idea that one may illuminate the meaning of normative vocabulary in terms of their communicative role:

It is worth mentioning that ethical terms do not serve only to express feeling. They are calculated also to arouse feeling, and so to stimulate action. [...] In fact we may define the meaning of the various ethical words in terms both of the different feelings they are ordinarily taken to express, and also the different responses which they are calculated to provoke. (p. 111)

If what I am about to see here is right, Ayer is a bit too nonchalant when he says that there is not much to choose between focussing on expression versus coordination in analyzing

normative meaning. There should be no doubt, however, that taking coordination as guide to the meaning of normative terms is both in the spirit as well in the letter of Ayer’s noncognitivist outlook on the language of morals.

None of this, for sure, is to say that the early noncognitivists are dynamic semanticists in disguise. It is perfectly legitimate to insist that the dynamic effects of moral language—just as any dynamic effects in discourse—are a merely pragmatic affair. We are also well-advised to take [Hare \(1952\)](#) seriously when he urges not to confuse *telling* someone that stealing is wrong with *making* him or her disapprove of stealing—a distinction that, if Hare is right, Ayer’s and Stevenson’s proposal fail to capture. The point here is that from the perspective of the early noncognitivists, it would be quite surprising if moral language use were all about expressing states of mind. In contrast, a dynamic interpretation—treating moral sentences as transition instructions on states of mind that differ from belief in that they are conative, desire-like states—is very much in line with the key intuitions that early noncognitivists have articulated about the language of morals. It remains to be seen, of course, that a dynamic proposal delivers the desirable results. Showing that this is the case is the purpose of the next sections.

3 The Dynamic Perspective

The purpose of the current exercise is to make a case for replacing expressivism with a dynamic implementation of the noncognitivist agenda in metaethics. §3.1 outlines the basic ideas and §3.2 elaborates the details.

3.1 Basics

The initial thought is familiar and begins with [Stalnaker’s \(1978\)](#) truisms about assertions. Every well-run conversation takes place against the background of a discourse context, that is, the common ground between the discourse participants—a state of information of a certain kind (rather than a concrete location). Assertions express propositions and since language contains indexicals, which proposition an assertion expresses depends on the context in which it is made. But assertions in turn affect the context, and they do so by adding the proposition expressed to the common ground. For Stalnaker, this effect is a purely pragmatic phenomenon, but his truisms about assertion may also serve as a source of inspiration for a change in perspective: instead of following the classical path by stating the semantic value of a declarative sentence in terms of its truth-conditions, we may ask how it relates an input context (the context in which it is uttered) to an output context (the context posterior to that utterance). Meanings thus become relational: they are relations between contexts.⁸

Thinking of semantic values as relations between contexts raises the question what a context is, and here we have a lot of room for interpretation. Putting the focus on discourse effects would suggest that we think of contexts as abstract representations of the

⁸My strategy here of moving from uncontroversial observations about the dynamics of conversation to a dynamic semantic perspective on meaning and communication owes inspiration to [Dever 2006](#) and [von Stechow and Gillies 2008](#). Classical dynamic semantic frameworks include: Discourse Representation Theory ([Kamp \(1981\)](#); [Kamp and Reyle \(1993\)](#); [Kamp et al. \(2011\)](#)), Dynamic Predicate Logic ([Groenendijk and Stokhof \(1991\)](#)), File Change Semantics ([Heim \(1982\)](#)), Update Semantics ([Veltman \(1985, 1996\)](#)).

common ground, but for current purposes the more salient option is to think of contexts as abstract representations of an agent’s state of mind. Doing so highlights the relation between the upcoming dynamic proposal and the expressivist semantics considered earlier: instead of thinking of language as expressing states of mind, we think of language as articulating *relations* between states of mind. Not surprisingly, the proposal then is that moral and descriptive sentences differ in that the former are designed to induce moral attitudes while the latter are designed to induce beliefs. To keep track of this distinction, I will sometimes speak of *noncognitive* and *cognitive* content and the idea is that the former is an update procedure on noncognitive (desire-like) states while the latter is an update procedure on cognitive states (beliefs and such).

To get a feel for how all of this might work out in detail, start with an idea that is straightforward to make precise: that descriptive sentences are designed to induce beliefs. Consider a simple propositional language \mathcal{L}_0 whose atomic sentences only involve descriptive predicates and model a state of belief in the usual fashion as a set of possible worlds: the set of possible worlds compatible with what the agent believes. On this model, an agent believes that ϕ just in case ϕ is true at every possible world compatible with what the agent believes, and semantic values are simply relations between belief states. Precisely, a possible world is (for current purposes) just a function from atomic sentences $\mathcal{A} = \{p, q, r, \dots\}$ to the truth-values 0 and 1. W is the set of all possible worlds, $\mathcal{P}(W)$ is the set of all sets of possible worlds, and a belief state $\sigma \in \mathcal{P}(W)$ is simply a set of possible worlds. The semantic value of a sentence ϕ , $[\phi]$ is a function $[\phi]: \mathcal{P}(W) \mapsto \mathcal{P}(W)$ from belief states to belief states, thus capturing the dynamic relational approach to semantics (following Heim (1983), who in turn builds on the proposal from Heim (1982)):

1. $\sigma[p] = \{w \in \sigma : w(p) = 1\}$
2. $\sigma[\neg\phi] = \sigma \setminus \sigma[\phi]$
3. $\sigma[\phi \wedge \psi] = \sigma[\phi][\psi]$

Update with an atomic sentence eliminates, from the input state, all possible worlds at which the atomic sentence is false. Negation amounts to eliminating from the input state the result of updating it with what is negated. Conjunction is just sequential update.

What is crucial about this semantics, for our purposes, is that it immediately predicts that an assertion that ‘Grass is green’ is designed to induce the belief that grass is green, while an assertion that ‘Grass is not green’ is designed to induce the belief that grass is not green. Moreover, any sentence and its negation are inconsistent in the sense that an update with ϕ followed by an update with $\neg\phi$ is guaranteed to result in the empty set. Insofar as our state of information is a representation of the state of belief, and assuming that belief is an inconsistency-transmitting attitude, we thus arrive at an explanation of the inconsistency between a sentence and its negation that has a familiar ring: updating any state with both sentences results in a state that instantiates two beliefs toward inconsistent descriptive contents, represented here by the empty set of possible worlds. And in general, whenever ϕ and ψ are guaranteed to induce two instances of an *A*-type inconsistent attitude toward inconsistent descriptive contents, we have an *A*-type explanation of why ϕ and ψ are inconsistent.

The suggestion then is that in order to arrive at a noncognitivist semantics for moral terms, we exploit the same strategy that we successfully applied in our model for the toy language \mathcal{L}_0 . Moral sentences are designed to induce some very general noncognitive

attitude, and I will follow Schroeder here in calling it the attitude of being for. And just as ‘Grass is green’ and ‘Grass is not green’ are inconsistent in the sense that they induce the same inconsistency-transmitting attitude toward inconsistent contents, so do ‘Stealing is wrong’ and ‘Stealing is not wrong.’ Schroeder does not say much about the attitude of being for, and nor will I, except that it is an inconsistency-transmitting attitude (toward action descriptions) that can motivate action, and I take it to be a strength of the upcoming proposal that it is compatible with a variety of outlooks on what exactly a state of being for would be. But elaborating upon the idea at play here requires a more complex model of an agent’s state of mind—one that distinguishes between what an agent believes and his or her state of being for. It also requires that our semantics distinguishes between the update effects of moral and descriptive sentences. Before that, let me briefly alleviate Hare’s worry that focussing on the role of moral sentences in coordinating attitudes of some kind conflates telling someone that stealing is wrong and making someone disapprove of stealing.

Thinking of semantic values dynamically as updates or transition instructions is not to say that assertions are instructions or commands. The reason is that settling on a semantics for declarative sentences still leaves work to do for a pragmatic theory. A natural thing to say is that an assertion of ϕ is commonly made with the intention that hearers update their information states with the content, that is the update potential, of ϕ . So an assertion of ‘Stealing is wrong’ is most naturally understood as a *proposal* to update and as such may be rejected by the hearer, and it may even be uttered with the expectation that the hearer *will* in fact reject the proposal. A speaker may, of course, react in various ways in case his or her assertion is rejected. Persuasion is one available option but there is, as far as the proposal here is concerned, no commitment to the claim speakers always choose this path. Hare’s distinction between telling and making someone accept what is told is perfectly robust in what I am about to say here.

3.2 Details

Earlier I outlined a dynamic toy model for a purely descriptive language in which sentences are designed to coordinate beliefs. Now we want to allow for the target language to also include moral vocabulary designed to coordinate on a distinguished kind of noncognitive attitude. To get the contrast between moral and descriptive predicates into better view, we let atomic sentences consist of n -place predicates and n constants.

Definition 1 (Language) The basic vocabulary of \mathcal{L} consists of n -place predicates F_1^n, F_2^n, \dots (for each $n \in \mathbb{N}$), constants $c_1, c_2 \dots$, the logical connectives \neg (negation) and \wedge (conjunction), and parentheses $(,)$. The set of well-formed formulas of \mathcal{L} is defined in the familiar fashion. Disjunction (\vee) and the material conditional (\supset) are defined in terms of negation and conjunction in the usual way. Subscripts and superscripts are suppressed whenever this does not give rise to confusion.

This setup does not draw a syntactic distinction between ‘Grass is green’ and ‘Stealing is wrong’—both sentences are of subject-predicate form—and so it is the job of our semantics to distinguish between them in a way that is sympathetic to the noncognitivist intuition that only the former have cognitive content. Omitting quantifiers and variables allows us to get the key ideas of the proposal into clear view.

Classical semantics assigns to sentences of the target language truth-values relative to a possible world. This will not do here since—at least for our purposes—moral sentences lack truth-conditions with respect to possible worlds: they are not accurate or inaccurate ascriptions of properties to objects and, relatedly, moral predicates lack extensions relative to a possible world. To streamline the proposal, we identify possible worlds with properly restricted first-order valuation functions: they assign to constants of the target language some world-invariant referent and to a distinguished set of predicates—those that count as descriptive—a suitable extension (which, of course, may vary across possible worlds).

Definition 2 (Possible Worlds) Given some underlying *domain* of individuals D , a possible world w maps (i) each constant c of \mathcal{L} to its world-invariant *denotation* $d \in D$ and (ii) a world-invariant subset of the n -place predicates of \mathcal{L} to some set of n -tuples of objects from D (for each $n \in \mathbb{N}$). W is the set of all possible worlds. A predicate of \mathcal{L} is *descriptive* iff it is in the domain of the elements of W .

The important bit to notice here is that certain predicate expressions—those we count as moral and thus as non-descriptive predicates—lack extensions at possible worlds (see also Gibbard (1990, 2003)).

The fact that certain predicate expressions lack extensions at possible worlds requires us to rethink the nature of terminal semantic values since sentences containing such expressions fail to have truth-values at possible worlds. The alternative I wish to pursue here is to model the semantics of sentences of \mathcal{L} in terms of update procedures on (abstract representations of) an agent’s state of mind. Doing so requires a precise conception of what such abstract representations should be. We already know how to model the content of an agent’s belief—as a set of possible worlds. For instance, $p \subseteq W$ represents grass as green iff $w(\mathbf{grass}) \in w(\mathbf{green})$ for all $w \in p$, and so a belief with p as its content is a belief that grass is green. What remains to be done is to add to this model a reasonable conception of the contents of those attitudes that moral sentences are designed to affect (attitudes of being for). The basic idea is that such attitudes have descriptions of actions as their content.⁹ For our purposes we simply identify actions with properties of agents—the agents performing those actions at possible worlds. Accordingly, a description of an action can simply be modeled as a set of agent-world pairs. For instance, if a is such a set, it represents the act of stealing just in case $x \in w(\mathbf{stealing})$ for all $\langle x, w \rangle \in a$, and so a state of being for with a as its content is a state of being for stealing.

One obvious path would then be to model an agent’s state of mind as two distinct sets of formal objects, one representing what an agent believes and the other representing what the agent is for. The problem with this approach is that the moral attitude under consideration should be sensitive to what is believed so that what one is for may depend on one’s beliefs about the actual world—hopes and desires are certainly belief-sensitive in this way.¹⁰ My desire to go to the party and my hope that Carl will come too,

⁹This approach to the objects of the relevant noncognitive attitude is in line with the path taken by Schroeder (2008a), who observes that it would do just as well for him to replace representations of actions with representations of facts—the same holds for the current proposal.

¹⁰The point is also observed by Heim (1990), Stalnaker (1984), and more recently by Silk (2015). It is compatible with the claim that the relevant attitudes are sometimes sensitive to features of an information state that arguably resist an adequate characterization as an attitude toward possible worlds, for instance to credences (as discussed by Cariani (2013, forthcoming) and Yalcin (2012), among others), though capturing such sensitivities would require a more complex model of an agent’s state of mind.

say, depends on my belief that the party will be fun and that Carl will not bring his annoying cousin along. So what I will do here is to pair each possible world in the belief set with a nonempty set of agent-world pairs (a representation of a possible complex action) capturing what the agent is for should that world turn out to be the actual world. Modeling a state of mind as a (possibly empty) set of such pairs gives us everything we need to keep track of an agent’s beliefs and of what the agent is for.

Definition 3 (States of Mind) Let $A = D_a \times W$, where $D_a \subseteq D$ is a set of *agents*. $\rho \in W \times (\mathcal{P}(A) \setminus \emptyset)$ is a *representation of a possible action α conditional on a possible world w* ; w_ρ and α_ρ are the possible world and possible action of ρ , respectively. R is the set of all ρ ’s. A *state of mind* σ is just a subset of R , and Σ is just the set of all states of mind. The *initial* state of mind σ_0 is identical with R , the *absurd* state of mind σ_\emptyset is identical with \emptyset .

A nonempty set of possible worlds represents a possible state of affairs, that is, whatever state of affairs holds at each of those possible world; similarly, a nonempty set of agent-world pairs represents a possible action, that is, whatever action is performed by each agent at every possible world. Coupling such a set with a possible world then represents what the agent is for given the facts in that world, and we treat a set of such conditional action representations as an abstract model of an agent’s state of mind.

It is then straightforward to identify what, given some state of mind, an agent is for and what the agent believes:

Definition 4 (Belief, Being For) Consider arbitrary $\sigma \in \Sigma$, $p \subseteq W$, and $a \subseteq A$:

1. $\sigma_{\text{Bel}} = \{v \in W : \exists \rho \in \sigma. v = w_\rho\}$
2. $\sigma_{\text{For}} = \{a \in A : \exists \rho \in \sigma. a \in \alpha_\rho\}$
3. σ is a state of *believing* the proposition p iff $\sigma_{\text{Bel}} \subseteq p$
4. σ is a state of *being for* the action a iff $\sigma_{\text{For}} \subseteq a$

States of belief and states of being for are, as explained earlier, treated as attitudes toward certain descriptive contents—descriptions of states of affairs and of actions, respectively. A subject believes that p is the case just in case all possible worlds of the subject’s belief state are p -worlds. A subject is for a just in case every agent-world pair $\langle x, w \rangle$ of the subject’s being for state is such that x performs a at w . Note here that since a state of mind may be absurd we leave room for an agent to believe an impossible proposition or to be for an impossible action.¹¹

It is then straightforward define cross-categorical notions of classical entailment (see [von Fintel \(1999\)](#)) for the contents of the belief and being for attitude. We also state precisely a generalized notion of negation for descriptive contents and define cross-categorical inconsistency on that basis:

¹¹How may we can account for the possibility of reflecting on what to do conditional on possibilities that are ruled out as candidates for being the actual world? By embellishing the model of an agent’s state of mind with a fallback relation that determines how σ_{Bel} would need to be weakened in the course of giving up on an existing rational commitment (see the discussion by [Willer \(2014\)](#) and references therein).

Definition 5 (Cross-Categorical Classical Entailment) Let the set of objects of type t consist of the True and the False and recursively define a cross-categorical relation of *classical entailment* (\Rightarrow) as follows:

1. For p, q of type t : $p \Rightarrow q$ iff $p = \text{False}$ or $q = \text{True}$.
2. For f, g of type $\tau \rightarrow \nu$: $f \Rightarrow g$ iff for all x of type τ , $\tau(x) \Rightarrow \nu(x)$.

Definition 6 (Cross-Categorical Classical Inconsistency) We say that that f and g are *inconsistent* iff $f \Rightarrow \sim g$, where:

1. If p is of type t , then $\sim p = \text{True}$ iff $p = \text{False}$.
2. If f is of type $\tau \rightarrow \nu$, then $g = \sim f$ iff $g(x) = \sim f(x)$ for all objects x of type τ .

Propositions are essentially functions from possible worlds to truth-values and thus we predict that entailment between propositions amounts to the subset relation while their inconsistency amounts to lack of a nonempty intersection. Relatedly, an action description entails another just in case the former is a subset of the latter, and they are inconsistent just in case their intersection is empty.¹²

We now have everything in place to define the update rules for \mathcal{L} . The procedure is straightforward: we first lay out how elements of \mathcal{L} update elements of an information state, and then define a transition instruction for information states on that basis. What is needed, of course, is some principled way for determining whether an atomic sentence is a means for coordinating on a state of belief or a state of being for. The answer depends on the interpretation of the predicate involved, which in turn is settled by appropriate lexical entries for the basic vocabulary of \mathcal{L} . In classical intensional semantics a lexicon just does for every predicate what we do for only some predicate expressions: assign an extension relative to every possible world. Given the partiality of our valuation functions (possible worlds) this will not do here, but it does not take much to find an adequate alternative.

The alternative pursued here is to treat predicate expressions as functions from n -tuples of objects from our domain of individuals D to updates on conditional action representations.

Definition 7 (Predicate Expressions) The *dynamic valuation function* $[[\cdot]]$ maps each n -place predicate of \mathcal{L} to a function from n -tuples of objects from D and conditional action representations ρ to subsets of α_ρ .

The more specific proposal is that a moral predicate such as ‘wrong’ strengthens an action representation so that it entails the act of blaming for the denotation of its argument. Descriptive predicates are modeled as tests: they check whether the denotation of the argument is in the classical predicate extension at the world under consideration, and either eliminate the conditional action representation or leave it untouched depending on the outcome of the test. Once the full framework is in place, we will see more clearly why

¹²A controversial feature of this setup is that belief and being for are closed under cross-categorical entailment. Furthermore, certain approaches to the semantics of indexicals—specifically those appealing to centered propositions—require a refinement of the notion of inconsistency if it is supposed to ground the notion of inconsistency between attitudes (taking inspiration from MacFarlane (2014), for instance). The ensuing problems are orthogonal to the issues pursued here.

this guarantees that atomic descriptive sentences induce a belief toward some proposition and affect a state of mind's being for attitudes only insofar as those are conditional on some proposition. Let me illustrate the setup of our lexicon by giving lexical entries for the \mathcal{L} -analogues of the natural language predicates ‘wrong’ and ‘green.’

$$\llbracket \text{wrong} \rrbracket = \lambda y \lambda \rho \lambda x \lambda w. \langle x, w \rangle \in \alpha_\rho \text{ and } \langle x, y \rangle \in w(\text{blame for})$$

$$\llbracket \text{green} \rrbracket = \lambda y \lambda \rho \lambda x \lambda w. \langle x, w \rangle \in \alpha_\rho \text{ and } y \in w_\rho(\text{green})$$

To wax technical for a moment: the semantic values for moral and descriptive predicates are of the same type, taking n -tuples of objects and some conditional action representation ρ as input and returning a subset of α_ρ as output. Moral predicates enforce that the agent stands in a distinguished descriptive relation toward the n -tuple of objects: in the case of ‘wrong,’ which is a one-place predicate, the relation of blaming for the denotation of the argument. So while moral predicates are not directly assigned extensions at possible worlds, they are tools for coordinating on the descriptive contents of conditional action representations. Descriptive predicates test whether the n -tuples of objects have a certain property at w_ρ : in the case of ‘green,’ which is a one-place predicate, the property of being green. As one can see from its lexical entry, the predicate returns α_ρ in case this is so, and otherwise the empty set.

As advertised earlier, we state how sentences of \mathcal{L} update conditional action representations and then define transition instructions on states of mind on that basis. Notice that the first upcoming definition once again adopts Heim’s (1983) proposal to treat negation as set subtraction and conjunction as sequential update.

Definition 8 (Updates on Elements of States of Mind) Define an update function $\uparrow: (\mathcal{L} \mapsto (R \mapsto (A \mapsto A)))$ according to the following recursive rules (here we let d_1, \dots, d_n be the denotations of constants c_1, \dots, c_n , respectively):

1. $\alpha \uparrow_\rho \Pi^n c_1 \dots c_n = \llbracket \Pi_n \rrbracket(d_1) \dots (d_n)(\rho)$
2. $\alpha \uparrow_\rho \neg \phi = \alpha \setminus (\alpha \uparrow_\rho \phi)$
3. $\alpha \uparrow_\rho (\phi \wedge \psi) = (\alpha \uparrow_\rho \phi) \uparrow_\rho \psi$

Definition 9 (Updates on States of Mind) Associate with each $\phi \in \mathcal{L}$ a function $[\phi]: \Sigma \mapsto \Sigma$ as follows:

$$\sigma[\phi] = \{\rho' \in R: \exists \rho \in \sigma. w_{\rho'} = w_\rho \text{ and } \alpha_{\rho'} = \alpha_\rho \uparrow_\rho \phi\}$$

Accordingly, updating a state of mind comes down to the following procedure: first, update the action representation of each of its elements; then gather all the resulting elements, leaving out those that come with an inconsistent action description. This yields the output state.

All of this is just to say that an atomic descriptive sentence such as ‘Grass is green’ eliminates all conditional action descriptions where grass fails to be green at the world parameter, thus inducing the belief in the proposition that grass is green (recall Definition 4). An atomic moral sentence such as ‘Stealing is wrong’ eliminates from all conditional action descriptions those world-agent pairs $\langle x, w \rangle$ whose agent x fails to blame for stealing

at w , thus inducing the attitude of being for the act of blaming for stealing (again, see Definition 4). The definition of negation guarantees that ‘Grass is not green’ induces the belief that grass is not green while ‘Stealing is not wrong’ induces the attitude of being for the act of not blaming for stealing. A conjunction of ‘Grass is green’ and ‘Stealing is wrong’ induces the belief that grass is green and the attitude of being for the act of blaming for stealing—a fact to which we shall return momentarily.

The final step on the to do list is to define the notions of acceptance and entailment. The notion of consistency follows immediately from the definition of entailment but is stated explicitly here to facilitate the discussion.

Definition 10 (Acceptance, Entailment, Consistency) Consider any $\sigma \in \Sigma$:

1. σ *accepts* ϕ , $\sigma \models \phi$, iff $\sigma[\phi] = \sigma$
2. ϕ_1, \dots, ϕ_n *entails* ψ , $\phi_1, \dots, \phi_n \models \psi$, iff for all $\sigma \in \Sigma$: $\sigma[\phi_1] \dots [\phi_n] \models \psi$
3. ϕ_1, \dots, ϕ_n is *consistent* iff for some $\sigma \in \Sigma$: $\sigma[\phi_1] \dots [\phi_n] \neq \emptyset$

A state of mind accepts ϕ just in case updating σ with ϕ idles. An argument is valid just in case the result of updating any state of mind accepts its conclusion. Accordingly, a sequence is consistent just in case we can find some state of mind that may be updated with that sequence without resulting in the absurd state. The job of the next section is to take a closer look at the predictions that this framework makes.

4 Output

Return to the observation that (1) and (2) are inconsistent (repeated):

- (1) Stealing is wrong.
- (2) Stealing is not wrong.

Schroeder (2008a, 2008b)-style expressivism explains the inconsistency between (1) and (2) by saying that these sentences express the same inconsistency-transmitting attitude of being for toward inconsistent descriptive contents—just as ‘Grass is green’ and ‘Grass is not green’ express the same inconsistency-transmitting attitude of belief toward inconsistent descriptive contents—and it is straightforward to see how this idea is translated into the dynamic framework developed here. For suppose that Π^n is a descriptive predicate: then for some $p \subseteq W$, $\sigma[\Pi^n c_1 \dots c_n]$ is a state of believing that p while $\sigma[\neg \Pi^n c_1 \dots c_n]$ is a state of believing that $\sim p$. On the other hand, suppose that Π^n is a moral and hence a non-descriptive predicate: then for some $a \subseteq A$, $\sigma[\Pi^n c_1 \dots c_n]$ is a state of being for a while $\sigma[\neg \Pi^n c_1 \dots c_n]$ is a state of being for $\sim a$. To say that any state of believing p and $\sim p$, or of being for a and $\sim a$, is *absurd* is then just to say that the attitudes represented by our states of mind are inconsistency-transmitting: any two sentences that induce two instances of these attitudes toward inconsistent descriptive contents are inconsistent. Let me generalize this crucial moral of the story a bit.

I have explained negation in terms of ruling out thus: ‘ $\neg\phi$ ’ induces a state of ruling out accepting ϕ . I have also said that the relevant notion of ruling out is not a primitive one but amenable to an A -type explanation, and I can make this more precise now. The first general observation is that any two sentences of the form ϕ and ‘ $\neg\phi$ ’ from our target language are predicted to be inconsistent:

Fact 1 Consider arbitrary $\phi \in \mathcal{L}$: ϕ and $\ulcorner \neg\phi \urcorner$ are inconsistent.

This is just to say that whenever we have a state of mind $\sigma \in \Sigma$, then $\sigma[\neg\phi]$ is a state of ruling out accepting ϕ in the following sense: coming to accept ϕ in state $\sigma[\neg\phi]$ results in the absurd state, that is, $\sigma[\neg\phi][\phi] = \sigma_\emptyset$. This, of course, would not be much progress if adopting an absurd state of mind amounted to nothing more than adopting two primitive, logically unrelated attitudes that we can only stipulate to be inconsistent, but fortunately this is not the case here. The key observation is the following:

Fact 2 Consider arbitrary $\sigma \in \Sigma$ and $\phi \in \mathcal{L}$: $\sigma[\phi] = \sigma_\emptyset$ if and only if (i) for some $a \subseteq A$, $\sigma[\phi]$ is a state of being for a and a state of being for $\sim a$ or (ii) for some $p \subseteq W$, $\sigma[\phi]$ a state of believing p and a state of believing $\sim p$.¹³

As such, the notion of adopting an inconsistent state of mind amounts to nothing more than the one of having two instances of a single attitude toward inconsistent descriptive contents. Assuming that the attitude of being for is inconsistency transmitting like the one of belief, the dynamic inconsistency between (1) and (2) thus receives an A -type explanation: one that explains the inconsistency between (1) and (2) in terms of two instances of a single, inconsistency-transmitting attitude toward inconsistent descriptive contents.

The first good news, then, is that we can rewrite the most promising expressivist story about (1) and (2) with a dynamic spin. The even better news is that the dynamic rewrite takes the hurdles that its expressivist alternatives knocks over. The basic observation from §2, remember, is that (2) is not only inconsistent with (1) but also with (3), and so is what (4) says (repeated):

- (3) Stealing is wrong and grass is green.
- (4) Grass is not green.

As we saw earlier, the expressivist strategy of explaining logical inconsistencies as A -type inconsistencies yields the prediction that (2)–(4) all express the same kind of attitude, and this is unfortunate since belief now turns out to be a conative, desire-like state. The dynamic proposal avoids this consequence in just the right way. First, here is the relevant prediction:

Fact 3 Consider arbitrary $\phi, \psi \in \mathcal{L}$: ϕ and $\ulcorner \neg\phi \wedge \psi \urcorner$ are inconsistent.

And all of this without predicting that a conjunction such as (3) induces a single attitude toward some descriptive conjunctive content. Instead, it is easy to verify that (3) induces a state of being for with some descriptive content a and a state of belief with some descriptive content p . The inconsistency between (2) and (3) is then grounded in the fact that the former induces a state of being for with the descriptive content $\sim a$. Similarly, the inconsistency between (3) and (4) is grounded in the fact that the latter induces a

¹³The left-to-right direction is obvious. To see the right-to-left direction, suppose first that $\sigma[\phi]$ is a state of being for a and a state of being for $\sim a$: then $\sigma[\phi]_{\text{For}} = \emptyset$ and so $\alpha_\rho = \emptyset$ for all $\rho \in \sigma[\phi]$. But for all $\rho \in \sigma$, $\alpha_\rho \neq \emptyset$, hence $\sigma[\phi] = \sigma_\emptyset$. Suppose that $\sigma[\phi]$ is a state of believing p and a state of believing $\sim p$: then $\sigma[\phi]_{\text{Bel}} = \emptyset$ and hence $\sigma[\phi] = \sigma_\emptyset$ as well.

state of belief with the descriptive content $\sim p$. Updating any state with (2) and (3) or (3) and (4) thus results in the absurd state: the claim that two sentences are inconsistent just in case that they induce two instances of a single attitude toward inconsistent descriptive contents does not collapse into the claim that (3) induces a single attitude toward some *conjunctive* descriptive content.

The underlying observation here is that logical operators are not functions on the descriptive contents of certain attitudes but functions on update operations. This effectively allows us to say that, for instance, (3) induces a *conjunction* of attitudes rather than a single attitude toward some conjunctive content, the one being *A*-type inconsistent with the attitude induced by (2) while the other is *A*-type inconsistent with the attitude induced by (4). On first sight, dynamic negation might be understood as a function on the contents of attitudes since any atomic sentences $\Pi^n c_1 \dots c_n$ induces an attitude toward some descriptive content while $\neg \Pi^n c_1 \dots c_n$ induces that very same attitude toward the negation of that content. But this is not simply not so, as we can see when we consider the negation of (3):

(5) It is not the case that stealing is wrong and that grass is green.

Earlier we saw that negated conjunctions such as (5) pose a problem for expressivists if we think of conjunction as inducing combinations of attitudes rather than single attitudes toward conjunctive contents: negation can then now longer be analyzed as operating on the content of a single attitude, making Schroeder-style treatments of negation unavailable for our analysis. In contrast, negated conjunctions pose no trouble for the dynamic framework proposed here, which makes the following prediction:

Fact 4 Consider arbitrary $\sigma \in \Sigma$ and $\phi, \psi \in \mathcal{L}$: $\sigma[\neg(\phi \wedge \psi)] = \sigma[\neg\phi] \cup \sigma[\neg\psi]$

Such a state is aptly called ‘a state of ruling out the conjunction of ϕ and ψ ’ since we have $\sigma[\neg(\phi \wedge \psi)][\phi][\psi] = \sigma_\emptyset$ by design.

I will return to negated conjunctions momentarily when I state more precisely what is at stake in a dynamic response to the Frege-Geach problem. Before that, let me briefly point out here that, assuming the classical definition of disjunction in terms of conjunction and negation, the analysis of (5) is, *mutatis mutandis*, an analysis of (6):

(6) Stealing is wrong or grass is green.

Specifically, observe that $\sigma[\phi \vee \psi] = \sigma[\phi] \cup \sigma[\psi]$. This is important: [Schroeder \(2008a\)](#) details that expressivists are committed to the prediction that (6) expresses a state of being for blaming for stealing or believing that grass is green (see also [Schroeder 2015](#), Ch. 9 for detailed discussion). This is an unacceptable result since a commitment to a disjunction is not to be analyzed as a commitment to one of its disjuncts. The dynamic analysis avoids this prediction. As we can see, it makes sense to say that a disjunction such as (6) induces a conjunction of *conditional* states of mind: the state of being for *a* (for some $a \subseteq A$) under the supposition that grass is not green, and the state of believing that *p* (for some $p \subseteq W$) under the supposition that stealing is not wrong (and relatedly for negated conjunctions). This strikes me as exactly the right result.

Moral and descriptive sentences figure in logical arguments in the familiar fashion. The previous observation about disjunction together with the classical definition of the

material conditional, for instance, immediately predict that modus ponens is valid in the following sense:

Fact 5 $\phi \supset \psi, \phi \models \psi$

All of this takes care of Geach-style worries that noncognitivists have no good explanation for why moral thought and talk enters into the same inferential relations as does descriptive thought and talk.

A dynamic perspective on language and communication allows us to synthesize ideas from early and contemporary metaethical noncognitivism to make real progress toward a satisfying perspective on moral and descriptive language. From the more recent literature we take the insight that logical inconsistency is to be explained in terms of *A*-type inconsistency between states of mind, and a closer look at what early noncognitivists have to say about the language of morals motivates a dynamic implementation of this idea that looks at the potential of language for *coordinating*—rather than merely expressing—states of mind. Once we have arrived at a sufficiently rich model of such states of mind that distinguishes between doxastic and non-doxastic attitudes toward descriptive contents, it is straightforward to arrive at a semantics that not only resolves the classical noncognitivist problems with negation but also more complex issues affecting Schroeder-style expressivism, such as those surrounding mixed conjunctions and disjunctions.

The technical apparatus outlined here may be expanded and modified in several ways. Most obviously, we should take a closer look at what happens once we take quantifiers into account (including tenses and modals), but doing so goes beyond of the scope of this discussion and is better reserved for another day. However, the success of the framework developed here together with the fact that there are well-established methods for handling quantification in dynamic semantics (see in particular [Groenendijk et al. 1996](#) and [Beaver 2001](#)) should establish some optimism that doing so will not lead to any bad surprises. So let me conclude the discussion by taking a closer look at what exactly is non-negotiable about the dynamic story told here and how it relates to some more recent trends in the expressivist literature.

5 Dynamics?

I have already explained why the the dynamic story told here offers an attractive alternative to the classical expressivist proposals from [Blackburn \(1984, 1988\)](#) and [Gibbard \(1990, 2003\)](#) and to the expressivist path pursued, but ultimately found to be unworkable, by [Schroeder \(2008a, 2008b\)](#). The purpose of this section is to compare the current proposal with a number of accounts from the quite recent expressivist literature. The point of all this is to give these proposals their proper due but also to highlight some distinct contributions that the story told here has to offer.

Several authors have recently suggested that expressivists (or anyone who is sympathetic to the noncognitivist agenda in metaethics) may help themselves to a truth-conditional semantics and interpret expressivism as a pragmatic thesis about language use.¹⁴ The semantic suggestion, in brief, is that ‘Stealing is wrong’ and ‘Stealing is not

¹⁴See the discussions by [Silk \(2013\)](#), who assigns to moral sentences truth-values relative to world-standard pairs, and [Charlow \(2014\)](#), who is less explicit about the semantic details but argues that

wrong’ are inconsistent in virtue of their imposing incompatible constraints on a *non-factual* parameter: taking the story from Gibbard (1990, 2003) as a starting point, for instance, a world-norm pair $\langle w, \pi \rangle$ is no longer interpreted as a model of an agent’s state of mind but rather as an abstract index of evaluation. Semantic values are now set of world-norm pairs, where descriptive sentences impose constraints on the world-parameter while moral sentences impose constraints on the norm-parameter. The inconsistency between ‘Stealing is wrong’ and ‘Stealing is not wrong’ is then no more surprising than the one between ‘Grass is green’ and ‘Grass is not green’—all there is to it that each sentence and its negation have disjoint semantic values by design—and nor is the potential of moral language to figure in logically valid arguments.

All of this talk about moral declarative sentences having truth-conditions, so the story continues, is consistent with a nonrepresentational *use* of moral language. Specifically, it is open for expressivists to maintain that moral sentences encode a nonrepresentational perspective in virtue of their imposing constraints on the nonfactual parameter of an index of evaluation, and say that moral judgments are proposals to psychologically approximate this nonrepresentational perspective (say, by adopting a state of planning in accordance with the norms represented by the relevant moral judgment’s semantic value). Descriptive sentences encode a representational perspective in virtue of their imposing constraints on the factual parameter of an index of evaluation, and descriptive judgments are proposals to psychologically approximate this representational perspective (say, by adopting a belief consistent with the set of possible worlds represented by the relevant descriptive judgment’s semantic value). Expressivism, on this view, is not tied to a distinct semantic thesis but rather focusses on the unique pragmatics of the language of morals.

The strategy I have outlined in the previous paragraph has a clear dynamic ring and is thus sympathetic to what I have said before, but it makes dynamic effects—and, more generally, the core commitments of expressivism—a matter of pragmatics while preserving a truth-conditional perspective on semantics.¹⁵ Why, then, choose a dynamic *semantic* path rather than the more classical one that treats dynamics as a pragmatic affair? The answer, in brief, is that pragmatic expressivism at best delays—but does not avoid—the key explanatory challenge that its semantic cousin faces: the need to explain why the states of mind expressed by ‘Stealing is wrong’ and ‘Stealing is not wrong’ are inconsistent. For sure, meeting this explanatory challenge is not required to get the semantics off the ground, since semantic values are now understood as sets of abstract indices of evaluation rather than as models of a more fundamental psychological reality. But every moral semantics worth its salt *eventually* needs to explain why the state of accepting that stealing is wrong is inconsistent with the state of accepting that stealing is not wrong, and it is anything but clear what pragmatic expressivism has to offer here.

truth-conditional semantics is compatible with a variety of ‘nonpropositionalist’ positions in philosophy of language, including metaethical expressivism. Yalcin (2007, 2011, 2012) develops an expressivist theory about the language of modality and probability that, so he maintains, lives happily with a truth-conditional semantics that assigns to such sentences truth-conditions relative to a possible world and some additional parameters.

¹⁵The idea that metaethical expressivism is just a pragmatic affair that may help itself to some notion truth at the level of semantics is related to what Horwich (1993, 1994) proposes when he says that expressivists should rely on a minimalist conception of truth in addressing their embedding problem, and Stoljar (1993) explicitly couples this suggestion with a nonrepresentational pragmatic story about moral language use. Horwich’s and Stoljar’s proposals are critically discussed by Dreier (1996).

Classical semantics explains the inconsistency between the belief that grass is green and the one that grass is not green in terms of the inconsistency of its contents: they are incompatible *representations* of a possible world in the sense that one of them must be accurate and the other be inaccurate. Such an explanation is not available for pragmatic expressivists when it comes to explaining the inconsistency between the state of accepting that stealing is wrong and the one of accepting that stealing is not wrong, even if we say that their contents impose conflicting constraints on some nonfactual parameter. Such states are not representational (any more than uses of ‘Stealing is wrong’ and ‘Stealing is not wrong’ are) and thus do not represent or misrepresent a norm, preference, or plan. Their inconsistency, accordingly, must be grounded in nonrepresentational aspects of the respective states—a conflict in plan, desire, preference, or what have you—and it is unclear why pragmatic expressivists should have a better answer to the question of *what* this aspect should be than their semantic cousins. We may, of course, say that, in virtue of their inconsistent contents, the states approximate or accord with conflicting norms, but so much was already predicted by the hyperplan model from Gibbard (1990, 2003). In other words, it is unclear how assigning to ‘Stealing is wrong’ and ‘Stealing is not wrong’ inconsistent but nonfactual contents is supposed to help explaining why the states of mind expressed by these sentences are inconsistent.

To get the worry into better perspective, consider the pragmatic expressivist outlook on the language of modality as developed by Yalcin (2007, 2011, 2012). Suppose that for any agent a and possible world w , there corresponds the set of worlds $B(a, w)$ compatible with the way a believes the world to be in w and the set of hyperplans $H(a, w)$ left open by a ’s planning state of mind at w . Suppose also that the modal epistemic *might* (\Diamond_e) and the modal deontic *ought* (\Box_d) receive truth-conditions relative to a possible world and nonfactual parameters s (a state of information understood as a set of possible worlds) and h (a hyperplan understood as a function from an information state to a set of possible worlds) that ‘float free’ of the world of evaluation w :

1. $\llbracket \Diamond_e \phi \rrbracket^{w,h,s} = 1$ iff $\exists w' \in s: \llbracket \phi \rrbracket^{w',h,s} = 1$
2. $\llbracket \Box_d \phi \rrbracket^{w,h,s} = 1$ iff $\forall w' \in h(s): \llbracket \phi \rrbracket^{w',h,s} = 1$
3. a accepts ϕ at world w iff $\forall h' \in H(a, w): \forall w' \in B(a, w): \llbracket \phi \rrbracket^{w',h',B(w,a)} = 1$

Here the first entry displays the truth-conditions of the epistemic possibility operator: it requires that the prejacent is compatible with the information carried by s . The second entry displays the truth-conditions of the deontic necessity operator: the prejacent is entailed by what the hyperplan h prescribes in light of s . The final entry defines the notion of acceptance in an obvious way.

Two observations about this setup are relevant for our purposes. First, it follows immediately that $B(a, w) = \emptyset$ whenever an agent a accepts a *might*-claim and its negation at some possible world w . Hence a ’s belief state at w leaves no possible world open, and this kind of inconsistency receives a straightforward *A*-type explanation as involving two instances of the attitude of belief toward inconsistent descriptive contents. So a non-factualist proposal for epistemic modals has no trouble explaining why, say, the state of accepting that it might be raining is incompatible with the one of accepting that it can’t be raining.

The second observation is that the case looks different if we consider an agent a who accepts an *ought*-claim and its negation at some possible world w . We immediately predict

that $H(a, w) = \emptyset$, which is just to say that the agent’s planning state leaves no hyperplan open. But a planning state is not—without additional arguments anyway—a description of a plan and so its failing to leave some hyperplan open is not an inconsistency in descriptive content. So while it is easy enough to explain why a belief state that rules out every possible world is inconsistent by appealing to familiar features of its representational content, this kind of explanation does not immediately apply when we try to understand what is wrong with a planning state that rules out every possible hyperplan. We are of course free to say that such planning states are in some sense inconsistent, but the pragmatic expressivist setup under consideration does not immediately dispel the worry that the inconsistency at play here is a primitive one and distinct from the familiar inconsistency in belief.

It may very well be that truth-conditional semantics can inform an account of the inconsistency between stealing is wrong and stealing is not wrong, but certainly the latter does not immediately fall out of the former. In contrast, the dynamic semantic framework developed in the previous sections explains why the state of accepting that stealing is wrong is inconsistent with the state of accepting that stealing is not wrong: these are attitudes toward inconsistent contents in the sense that no one could come to accept both without adopting an *A*-type inconsistent state of mind. It is thus in virtue of demonstrating that acceptance of inconsistent nonclassical semantic contents is amenable to an *A*-type explanation in the moral domain that the story told here offers a real explanatory advantage over pragmatic expressivism.

To be clear, the claim here is not that every explanation of inconsistency between noncognitive states of mind must be an *A*-type explanation. Building on [Dreier’s \(2006\)](#) account, [Silk \(2015\)](#) suggests that there is an alternative to this strategy that also avoids appealing to stipulated inconsistencies between primitive, logically unrelated attitudes (*B*-type inconsistencies). On his proposal, accepting that stealing is wrong and that stealing is not wrong involve incoherent preferences, where such preferences are non-representational features of a state of mind. A detailed assessment of this strategy is better left to another day (see [Starr forthcoming](#) for critical discussion), but let me make two brief remarks about how it connects with the current discussion. First, if successful, the strategy leads to a stable version of semantic expressivism: we may now simply identify the semantic value of a sentence with the state of mind it conventionally expresses, and explain the inconsistency between a moral sentence and its negation in terms of a practical incoherence between the preferences of the states those sentences conventionally express. Truth-conditions, coupled with an expressivist pragmatics to account for nonrepresentational language use, are no longer needed to overcome the Frege-Geach problem. The point generalizes: once an expressivist can explain why accepting that stealing is wrong and that stealing is not wrong are inconsistent—and such an explanation needs to be given anyway—there is no need to choose pragmatic over semantic expressivism. And second, whatever the vices and virtues of this strategy, it does not rely on a generalization of the familiar *A*-type explanation of inconsistency between beliefs but rather appeals to a nonstandard conception of inconsistency between states of mind that includes the notion of practical incoherence. It may very well be that ‘*C*-type inconsistencies’ of this kind can offer a stable foundation for logical inconsistencies between sentences, but the strategy of building one’s semantics on its basis comes with its very own explanatory challenges and clearly differs from the path taken here.

I have explained why the dynamic framework differs from—and arguably has explanatory advantages over—pragmatic expressivism and any version of semantic expressivism that appeals to *C*-type inconsistencies. But one may still wonder whether there is a *principle* reason to choose a dynamic semantic path and, in particular, whether the proposal made here could not be rewritten so that it turns out to be properly expressivist after all. The question in fact relates to the familiar issue in the empirical semantic literature of whether a given natural language phenomenon really requires a dynamic semantic analysis or can be captured by a more classical semantic approach coupled with a suitably flexible pragmatics. Readers acquainted with this literature may find the issue particularly pressing for the project at hand due to the observation that the story told here, while formally dynamic, fails to make any interestingly dynamic predictions about the language of morals. For instance, our update function is idempotent and commutative in the following sense:

Fact 6 Consider arbitrary $\phi, \psi \in \mathcal{L}$ and $\sigma \in \Sigma$: $\sigma[\phi][\phi] = \sigma[\phi]$ (*idempotence*) and $\sigma[\phi][\psi] = \sigma[\psi][\phi]$ (*commutativity*).

Commutativity failures, in particular, are a hallmark of classical dynamic semantic frameworks that one finds in the literature—absent such hallmarks, the need for a *dynamic* semantic story about the language of morals is anything but obvious.

It strikes me as uncontroversial that the language of morals—at least as far it has been considered here—offers no *empirical* evidence in favor of any nonclassical approach to semantic theorizing. The best such an approach can do here is to match the predictions of a classical and static truth-conditional analysis, and as such the absence of genuinely dynamic forecasts in the framework developed here is a welcome result. All of this, however, is compatible with there being a *conceptual* motivation for the dynamic story I have told, and I have already explained how it improves upon the semantic expressivist agenda in metaethics: by offering an *A*-type explanation of logical inconsistency while preserving a robust distinction between cognitive and noncognitive content (captured here as a distinction between the kinds of states on which moral and descriptive sentences coordinate).

What is true is that one can derive from our dynamic analysis a semantics for our target language \mathcal{L} that has a distinctly expressivist flavor. Specifically, we may model the semantic value of a sentence as the states of mind accepting that sentence as follows:

$$\text{For all } \phi \in \mathcal{L}, \llbracket \phi \rrbracket = \{ \sigma \in \Sigma : \sigma \models \phi \}$$

It is even possible to identify, among the states of mind accepting some sentence ϕ , a single weakest member and let this state, rather than sets of the states accepting ϕ , serve as the semantic value of ϕ . Moreover, we have already seen that $\llbracket \Pi^n c_1 \dots c_n \rrbracket$ and $\llbracket \neg \Pi^n c_1 \dots c_n \rrbracket$ represent states of mind with inconsistent descriptive contents and so one might really wonder what gets lost if we choose *this* expressivism-friendly approach to semantic theorizing over the advocated dynamic analysis.

There is something clearly right about this proposal: on the dynamic analysis elaborated here, we may continue to say that sentences of \mathcal{L} are conventionally used to express states of mind, and we may identify such states with the ones that accept the sentence under consideration (or some distinguished member of the set of these states). However,

while assigning to elements of \mathcal{L} semantic values on this basis may very well be empirically adequate, its explanatory power essentially derives from the underlying dynamic analysis of sentential content as a transition instruction on states of mind. Let me explain.

Go back to the story told by Gibbard (1990, 2003): that story, recall, faced the problem that negation cannot be understood as the mere absence of the attitude expressed by what is negated. The expressivist proposal currently under consideration does not face the same problem since it predicts the state of mind of, say, accepting that stealing is wrong to be strictly stronger than the one of failing to accept that stealing is wrong. There remains the question, however, of what *justifies* the analysis of negation—understood as a function from states of mind to states of mind—in the current expressivist setting. Specifically, negation cannot, as in the classical fashion, simply be understood as set complementation since $\llbracket \neg\phi \rrbracket \neq \Sigma \setminus \llbracket \phi \rrbracket$ (recall that Σ is the set of all states of mind), and it will clearly not do to simply stipulate that negation maps a state to a suitable *A*-type inconsistent state of mind. It is not simply an operator on the content of the input output—simply recall negated conjunctions—and treating it as an operator on the attitude itself brings back the familiar question of what it could mean to negate an attitude (rather than its content). But without an independently motivated story about expressivist negation, its justification can only be derived from the underlying dynamic analysis of negation understood as a function on transition instructions. Accordingly, the possibility of rewriting the dynamic proposal in an expressivism-friendly fashion changes nothing about the fact that taking coordination rather than expression as the guide to semantic theorizing proves to be foundational to the progress we have made toward resolving the Frege-Geach problem for metaethical noncognitivism.

There is, in brief, nothing wrong with continuing to talk about sentences expressing states of mind. This is good news since such talk is certainly intuitive, and it does not make the dynamic analysis superfluous here since any expressivist analysis derived from it must ultimately appeal to its underlying dynamic apparatus in explaining negation and inconsistency. A dynamic analysis of the language of morals may very be unable to claim empirical advantages over its static alternatives, but this is compatible with it being able to add something substantial to the resolution of certain key conceptual issues with the noncognitivist agenda in metaethics.¹⁶

The highlight of the story told here is that taking coordination instead of expression as the starting point of semantic theorizing allows ethical noncognitivist to make genuine progress toward resolving the Frege-Geach problem. Negation, conjunction, and other logical connectives receive a natural interpretation that gives rise to an attractive explanation of logical inconsistency: two sentences ϕ and ψ are inconsistent just in case they together induce an absurd state of mind, where the underlying notion of an absurd state of mind receives an *A*-type explanation as one incorporating two instances of the attitude belief, or of the one of being for, toward classically inconsistent descriptive contents.

¹⁶Rothschild and Yalcin (forthcoming) suggest that every dynamic system satisfying idempotence and commutativity are amenable to a static rewrite in which ‘propositions’ understood as sets of points of evaluation mediate each and every context change—in fact, the mediating propositions may just be the states of mind of accepting the sentence initiating that change. We can remain neutral on the question whether Yalcin and Rothschild characterize the scope (as well as the limits) of static semantics correctly. The point here is that results like these are compatible with dynamic semantics playing an explanatorily more foundational role than its static rewrite in semantic theorizing: the relevant static notion of negation, for instance, may only be justifiable by appealing to the underlying dynamic semantics from which the static framework is derived.

The framework thus developed draws on insights from the classical expressivist literature while being in the spirit as well as the letter of the earlier noncognitivist perspectives on the language of morals, and it comes with some unique explanatory features that cannot be mimicked by frameworks treating expressivism as a pragmatic thesis about moral language use, even if the pragmatics itself has a distinctly dynamic flavor. Expressivism may very well have its comeback since it has, as we have seen, the option of going beyond the unhappy choice between *A*-type and *B*-type explanations of inconsistency. For now, however, those with general sympathies for metaethical noncognitivism may take heart in the fact that taking coordination rather than expression as the guide to semantic theorizing makes *A*-type explanations of inconsistency a viable option after all.

Acknowledgments Thanks to audiences at the 2014 Barcelona Workshop on Semantic Content and Conversational Dynamics (organized by Max Kölbel) and at the 2015 Central APA for helpful comments and discussion, and special thanks to Thony Gillies, Alex Silk, and an anonymous reviewer for written feedback on earlier versions of this paper that proved to be invaluable in preparing the final manuscript (the usual disclaimers apply). Support from the Franke Center for the Humanities is gratefully acknowledged.

References

- Alwood, Andrew. Forthcoming. “Non-descriptive Negation for Normative Sentences.” *Philosophical Quarterly*. <http://dx.doi.org/10.1093/pq/pqv067>.
- Ayer, Alfred Jules. 1936. *Language, Truth, and Logic*. London: Gollancz.
- Baker, Derek, and Jack Woods. 2015. “How Expressivists Can and Should Explain Inconsistency.” *Ethics* 125(2): 391–424. <http://dx.doi.org/10.1086/678371>.
- Beaver, David I. 2001. *Presupposition and Assertion in Dynamic Semantics*. Stanford: CSLI Publications.
- Blackburn, Simon. 1984. *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Clarendon Press.
- . 1988. “Attitudes and Contents.” *Ethics* 98(3): 501–517. <http://dx.doi.org/10.1086/292968>.
- Cariani, Fabrizio. 2013. “Epistemic and Deontic *Should*.” *Thought* 2(1): 73–84. <http://dx.doi.org/10.1002/tht3.61>.
- . Forthcoming. “Deontic Modals and Probabilities: One Theory to Rule Them All?” In *Deontic Modality*, ed. Nate Charlow and Matthew Chrisman. Oxford University Press.
- Charlow, Nate. 2014. “The Problem with the Frege-Geach Problem.” *Philosophical Studies* 167(3): 635–665. <http://dx.doi.org/10.1007/s11098-013-0119-5>.
- . 2015. “Prospects for an Expressivist Theory of Meaning.” *Philosophers’ Imprint* 15(23): 1–43. <http://dx.doi.org/2027/spo.3521354.0015.023>.

- Ciardelli, Ivano, Jeroen Groenendijk, and Floris Roelofsen. 2009. "Attention! *Might* in Inquisitive Semantics." *Proceedings of SALT XIX* 91–108.
- Dever, Josh. 2006. "Living the Life Aquatic: Does Presupposition Accommodation Mandate Dynamic Semantics?" <http://www.pragmatics.osu.edu/links/events/aquatic.pdf>. Manuscript, University of Texas at Austin.
- Dreier, Jamie. 1996. "Expressivist Embeddings and Minimalist Truth." *Philosophical Studies* 83(1): 29–51. <http://dx.doi.org/10.1007/BF00372434>.
- . 2006. "Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution." In *Oxford Studies in Metaethics*, ed. Russ Shafer-Landau, vol. 15, 217–233. New York: Oxford University Press.
- . 2009. "Relativism (and Expressivism) and the Problem of Disagreement." *Philosophical Perspectives* 23(1): 79–110. <http://dx.doi.org/10.1111/j.1520-8583.2009.00162.x>.
- von Fintel, Kai. 1999. "NPI Licensing, Strawson Entailment, and Context Dependency." *Journal of Semantics* 16(2): 97–148. <http://dx.doi.org/10.1093/jos/16.2.97>.
- von Fintel, Kai, and Anthony S. Gillies. 2008. "An Opinionated Guide to Epistemic Modality." In *Oxford Studies in Epistemology 2*, ed. Tamar Gendler and John Hawthorne, 32–62. New York: Oxford University Press.
- Frege, Gottlob. 1919. "Die Verneinung. Eine Logische Untersuchung." *Beiträge zur Philosophie des Deutschen Idealismus* I(3-4): 143–157.
- Geach, Peter T. 1960. "Ascriptivism." *Philosophical Review* 69(2): 221–225. <http://dx.doi.org/10.2307/2183506>.
- . 1965. "Assertion." *Philosophical Review* 74(4): 449–465. <http://dx.doi.org/10.2307/2183123>.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.
- . 2003. *Thinking How To Live*. Cambridge, MA: Harvard University Press.
- . 2013. *Meaning and Normativity*. New York: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199646074.001.0001>.
- Groenendijk, Jeroen, and Martin Stokhof. 1991. "Dynamic Predicate Logic." *Linguistics and Philosophy* 14(1): 39–100. <http://dx.doi.org/10.1007/BF00628304>.
- Groenendijk, Jeroen, Martin Stokhof, and Frank Veltman. 1996. "Coreference and Modality." In *The Handbook of Contemporary Semantic Theory*, ed. Shalom Lappin, 179–214. Cambridge, MA: Blackwell.
- Hare, Richard M. 1952. *The Language of Morals*. Oxford: Clarendon Press.
- Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts, Amherst.

- . 1983. “On the Projection Problem for Presuppositions.” In *Proceedings of the Second Annual West Coast Conference on Formal Linguistics*, ed. Michael Barlow, Daniel Flickinger, and Michael Westcoat, 114–126. Stanford: CSLI Publications.
- . 1990. “E-Type Pronouns and Donkey Anaphora.” *Linguistics and Philosophy* 13(2): 137–177. <http://dx.doi.org/10.1007/BF00630732>.
- Horgan, Terry, and Mark Timmons. 2006. “Cognitivist Expressivism.” In *Metaethics after Moore*, ed. Terry Horgan and Mark Timmons, 255–298. New York: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199269914.003.0013>.
- Horwich, Paul. 1993. “Gibbard’s Theory of Norms.” *Philosophy & Public Affairs* 22(1): 67–78. <http://dx.doi.org/10.1093/0199251266.003.0009>.
- . 1994. “The Essence of Expressivism.” *Analysis* 54(1): 19–20. <http://dx.doi.org/10.2307/3328098>.
- Kamp, Hans. 1981. “A Theory of Truth and Representation.” In *Formal Methods in the Study of Language, Part I*, ed. Jeroen Groenendijk, Theo Janssen, and Martin Stokhof, 277–320. Amsterdam: Mathematisch Centrum.
- Kamp, Hans, Josef van Genabith, and Uwe Reyle. 2011. “Discourse Representation Theory.” In *Handbook of Philosophical Logic*, 2nd ed., vol. 15, ed. Dov M. Gabbay and Franz Guenther, 125–394. Dordrecht: Springer. http://dx.doi.org/10.1007/978-94-007-0485-5_3.
- Kamp, Hans, and Uwe Reyle. 1993. *From Discourse to the Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer Academic Press. <http://dx.doi.org/10.1007/978-94-017-1616-1>.
- MacFarlane, John. 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. New York: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199682751.001.0001>.
- Portner, Paul. Forthcoming. “Imperatives.” In *The Cambridge Handbook of Formal Semantics*, ed. Maria Aloni and Paul Dekker. Cambridge University Press.
- Roelofsen, Floris. 2013. “A Bare Bone Attentive Semantics for *Might*.” In *The Dynamic, Inquisitive, and Visionary Life of ϕ , $?\phi$, and $\diamond\phi$* , ed. Maria Aloni, Michael Franke, and Floris Roelofsen, 190–215. Amsterdam: ILLC Publications.
- Rosen, Gideon. 1998. “Blackburn’s *Essays in Quasi-Realism*.” *Noûs* 32(3): 386–405. <http://dx.doi.org/10.1111/0029-4624.00106>.
- Rothschild, Daniel, and Seth Yalcin. Forthcoming. “On the Dynamics of Conversation.” *Noûs*. <http://dx.doi.org/10.1111/nous.12121>.
- Schroeder, Mark A. 2008a. *Being For: Evaluating The Semantic Program of Expressivism*. New York: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199534654.001.0001>.

- . 2008b. “How Expressivists Can and Should Solve their Problem with Negation.” *Noûs* 42(4): 573–599. <http://dx.doi.org/10.1111/j.1468-0068.2008.00693.x>.
- . 2008c. “What is the Frege-Geach Problem?” *Philosophy Compass* 3(4): 703–720. <http://dx.doi.org/10.1111/j.1747-9991.2008.00155.x>.
- . 2010. *Noncognitivism in Ethics*. New York: Routledge.
- . 2015. *Expressing our Attitudes*. Oxford: Oxford University Press.
- Schueler, G.F. 1988. “Modus Ponens and Moral Realism.” *Ethics* 98(3): 492–500. <http://dx.doi.org/10.1086/292967>.
- Schwartz, Jeremy, and Christopher Hom. 2015. “Why the Negation Problem is Not a Problem for Expressivism.” *Noûs* 49(4): 825–845. <http://dx.doi.org/10.1111/nous.12068>.
- Searle, John R. 1962. “Meaning and Speech Acts.” *Philosophical Review* 71(4): 423–432. <http://dx.doi.org/10.2307/2183455>.
- Silk, Alex. 2013. “Truth Conditions and the Meanings of Ethical Terms.” In *Oxford Studies in Metaethics, Volume 8*, ed. Russ Shafer-Landau, 195–222. New York: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199678044.003.0008>.
- . 2015. “How To Be an Ethical Expressivist.” *Philosophy and Phenomenological Research* 91: 47–81. <http://dx.doi.org/10.1111/phpr.12138>.
- Sinnott-Armstrong, Walter. 2000. “Expressivism and Embedding.” *Philosophy and Phenomenological Research* 61(3): 677–693. <http://dx.doi.org/10.2307/2653618>.
- Stalnaker, Robert C. 1978. “Assertion.” In *Syntax and Semantics 9*, ed. Peter Cole, 315–332. New York: New York Academic Press.
- . 1984. *Inquiry*. Cambridge, MA: MIT Press.
- Starr, William B. Forthcoming. “Dynamic Expressivism about Deontic Modality.” In *Deontic Modality*, ed. Matthew Chrisman and Nate Charlow. Oxford University Press.
- Stevenson, Charles Leslie. 1944. *Ethics and Language*. Oxford: Oxford University Press.
- . 1963. *Facts and Values*. Westport, CT: Greenwood Press.
- Stoljar, Daniel. 1993. “Emotivism and Truth Conditions.” *Philosophical Studies* 70(1): 81–101. <http://dx.doi.org/10.1007/BF00989663>.
- Unwin, Nicholas. 1999. “Quasi-Realism, Negation and the Frege-Geach Problem.” *Philosophical Quarterly* 49(196): 337–352. <http://dx.doi.org/10.1111/1467-9213.00146>.
- . 2001. “Norms and Negation: A Problem for Gibbard’s Logic.” *Philosophical Quarterly* 51(202): 60–75. <http://dx.doi.org/10.1111/1467-9213.00214>.
- van Roojen, Mark. 1996. “Expressivism and Irrationality.” *Philosophical Review* 105(3): 311–335. <http://dx.doi.org/10.2307/2185703>.

- Veltman, Frank. 1985. *Logics for Conditionals*. Ph.D. thesis, University of Amsterdam.
- . 1996. “Defaults in Update Semantics.” *Journal of Philosophical Logic* 25(3): 221–261. <http://dx.doi.org/10.1007/BF00248150>.
- Wedgwood, Ralph. 2007. *The Nature of Normativity*. New York: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199251315.001.0001>.
- Willer, Malte. 2014. “Dynamic Thoughts on Ifs and Oughts.” *Philosophers’ Imprint* 14(28): 1–30. <http://dx.doi.org/2027/spo.3521354.0014.028>.
- . 2015. “Simplifying Counterfactuals.” *Proceedings of the 20th Amsterdam Colloquium* 428–437.
- Yalcin, Seth. 2007. “Epistemic Modals.” *Mind* 116(464): 983–1026. <http://dx.doi.org/10.1093/mind/fzm983>.
- . 2011. “Nonfactualism about Epistemic Modals.” In *Epistemic Modality*, ed. Andy Egan and Brian Weatherson, 295–332. New York: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199591596.003.0011>.
- . 2012. “Bayesian Expressivism.” *Proceedings of the Aristotelian Society* 122(2pt2): 123–160. <http://dx.doi.org/10.1111/j.1467-9264.2012.00329.x>.