# 5

# Metaethical Expressivism

Elisabeth Camp

Expressivism is the view that certain kinds of language have the function of expressing states of mind rather than representing facts. So according to expressivists, when I say "Murder is wrong!" I don't describe a state of affairs, but avow or display or advocate a negative attitude toward murder. More specifically, expressivism holds that words like 'ought' or 'wrong' conventionally function to express *non-cognitive* attitudes: attitudes other than straightforward belief, such as emotions or intentions. It holds that these non-cognitive attitudes *explain* those words' meanings rather than just happening to be frequently correlated with their use. And it holds that the meaning and function of these words *differ* in a fundamental way from ordinary description. Different expressivists target different kinds of language, associate them with different attitudes, and locate the contrast with description in different ways, producing a diverse family of views.

Although expressivism is a view about linguistic meaning, it is natural to assume that language and psychology operate in parallel, especially if one takes the job of language to be communicating thoughts, as many do. As a result, expressivism is naturally allied to non-cognitivism, which is a view about the basic psychology of engagement with a topic, paradigmatically ethics. For both, the core idea is that we distort the shape of ethical inquiry, commitment, and disagreement if we treat ethical thought and talk in descriptivist terms, as a matter of exchanging information about how the world is. Metaphysically, a descriptivist model threatens to commit us to 'spooky', non-natural facts: abstract properties like *being wrong* that are unanchored to time, place, or particular social practices. Epistemically, it threatens to commit us to positing information whose discovery would or should resolve disputes, where many have thought that even total information about how things *are* still leaves open the question of what is right or wrong. And practically, a descriptivist model threatens to undercut apparently intimate connections between judgment and motivation: thus, it seems that if I think murder is wrong, I will or at least

should be motivated not to murder and to discourage murder, but it is unclear how bare facts could, by themselves, underwrite such motivation.

Instead, ethical non-cognitivists propose that the fundamental psychological states involved in considering what is right and wrong, or what to do, are desires, emotions, and/or intentions. Like beliefs, and unlike mere sensations or moods, these psychological attitudes are *about* more or less specific objects and situations: I desire, fear, hope for, and plan to bring about certain states of affairs. And like beliefs, they are related in at least somewhat systematic patterns of compatibility and inconsistency: if I fear that I will lose my job, there is at least *prima facie* something problematic about simultaneously planning to insult my boss to her face. But unlike beliefs, the function of such non-cognitive states is not to describe how the world *is*, but to show how the world *should* be, and to lead the agent to act accordingly. (For discussion, see Matt Bedke's chapter "Cognitivism and Non-Cognitivism".)

The contrasts between belief and these other attitudes, and between factual descriptions and avowals of feeling and intention, are fairly intuitive, as is the idea that ethical commitments involve feelings, preferences, and plans in some central way. The main challenge for the ethical expressivist is to ground these contrasts in the way we actually talk about ethics, which is typically with declarative sentences that behave a lot like ordinary assertions. Thus, the fact that we regularly say things like 'John believes that murder is wrong, and I agree, but I don't think it follows that abortion is wrong' has led many contemporary expressivists to grant that ethical statements do express beliefs, and are true, in some minimal sense. At the same time, some expressivists have also targeted domains like probability, epistemic modality, and knowledge, by appealing to modified versions of many of the same basic metaphysical, epistemic, and practical motivations as ethical expressivists. However, these expressivist analyses tend to be grounded in psychological states that are more belief-like, with functions that are more tightly tied to tracking and manipulating information. Many of these modifications to and extensions of 'classic' ethical expressivism are well motivated. But they also erode the initial, intuitive contrast between expressing attitudes and describing facts. This, together with the appropriation of belief- and truth-talk by many expressivists, renders the boundary between expressivism and its competitors increasingly blurry.

An additional complication arises from the fact that although non-cognitivism and expressivism are close cousins, they are distinct views about distinct subjects. In particular, a non-cognitivist could emphasize the central importance of feelings and plans in engaging with ethics while retaining a fundamentally descriptivist analysis of the words we use to talk about it. Such a theorist might think that simple sentences like 'Murder is wrong' are false because they ascribe properties that don't exist but take those false (or perhaps, pretended) claims to aptly reflect a broader non-cognitive psychology. Or they might hold that such sentences are potentially true but that their truth is relatively unimportant because speakers use them first and foremost to communicate non-cognitive states. Thus, we should view expressivism, of whatever specific form, as one commitment which fits together with non-cognitivism and anti-realism to form an especially elegant metaethical (Or metaepistemological, etc.) package, with each element potentially being leveraged in other combinations to form quite different views.

In this chapter, I examine expressivism as the claim that certain classes of words, especially 'thin' normative terms like 'ought' and 'good', conventionally function to express non-cognitive psychological states. In the next section, I consider what the relation between attitudes and utterances must be like to count as expressing rather than

describing. This is an issue in the general theory of meaning. In the section "How Do Words Express?", I turn to the more specific question of how language, as a conventional communicative system in which words combine to form whole sentences, might implement this relation. Throughout, we will find the expressivist striving to balance respect for the core intuition that normative language *does* something distinctive against the need to accommodate strong parallels between normative talk and straightforward description.

## WHAT IS EXPRESSING?

The ethical expressivist maintains that normative utterances function to communicate non-cognitive psychological states rather than to describe worldly states of affairs. However, it is not enough for those utterances to communicate those psychological states in just any way. In particular, the expressivist denies that those utterances communicate those states by *reporting* that the speaker has them, as 'I disapprove of murder' does. Such a simple subjectivist analysis would undermine the very contrast the expressivist wants to capture. More importantly, it radically misdescribes ethical discourse. When agents make claims about, provide reasons for, and challenge each other about ethical matters, they aren't arguing about their mental states. If they were, we wouldn't find even the appearance of disagreement. Instead, ethical utterances would articulate parallel but distinct and therefore compatible states, much as 'I am hungry' in your mouth and mine do. Further, it would be appropriate to evaluate those utterances as *true*, just in case the speaker did disapprove of murder.

How, then, should the expressivist understand the contrast with reporting, whether a worldly or a psychological state? One natural place to start is with the idea of "performative" utterances. In reaction to the dominant descriptivist model of language, John Austin (1961) drew attention to utterances like 'I declare you married', which don't represent an independent state of affairs as obtaining, and so are not straightforwardly truth-apt in the way that descriptions are. Instead, Austin characterizes their success conditions in terms of 'felicity': an assessment of whether the right social and psychological background conditions—such as being of sound mind, willing, of appropriate age, and not currently married to someone else—obtain. Similarly, the expressivist might plausibly claim that normative utterances serve to *do* something in conversation, rather than saying *that* something has been done, and so that they should be assessed for appropriateness rather than truth.

While this distinction gets something right, there are at least two problems with exploiting it to identify a distinctively expressive class of utterances. The first is that, as Austin argues, the distinction is not exclusive. On the one hand, descriptive statements are themselves performative, and subject to assessment on grounds besides truth: whether the speaker has good evidence for her claim, whether it is conversationally relevant or polite, whether it employs an apt classificatory scheme. And on the other hand, many performative utterances are evaluable in terms of "a general dimension of correspondence with fact" (Austin 1961, 250); thus, a call of 'foul ball!' is apt only if the baseball landed on the far side of the line. Given that many utterances are both performative and descriptive, the expressivist can't establish that canonical normative utterances *don't* describe by showing that they *do* accomplish something else. Recently, 'hybrid' expressivists like Michael Ridge (2006, 2014) have embraced this sort

of non-exclusivity, arguing that normative talk both describes and expresses (see Teemu Toppinen's chapter "Hybrid Accounts of Ethical Thought and Talk").

The second problem is that not just any connection between uttering a sentence and performing an action supports expressivism. The expressivist needs to identify, not merely a distinctive *doing* that people sometimes or often undertake in talking about ethics, but a conventional linguistic means for accomplishing it. We've already seen that subjective reports don't express in the relevant sense. But neither do utterances like 'Whenever I see a drowning baby, I rescue it', since they (at best) *implicate* the relevant non-cognitive attitude rather than communicating it directly. To establish expressivism as a claim about the function of language itself, rather than about the ways speakers exploit language in particular conversations, the expressivist needs to show that the uttered sentences' conventional role is the communication of non-cognitive attitudes. Dialectically, without evidence for a distinctive, conventional mechanism for expressing non-cognitive states, the expressivist has no argument against a standard semantic theory, with the communication of non-cognitive states being at most a pragmatic accompaniment to a conventional descriptive contribution.

The difference between direct and indirect modes of communication isn't just a matter of how meanings are produced; it also affects the role that utterances play in subsequent discourse. In particular, only content that is directly communicated and 'at-issue' (Potts 2005) is available for straightforward response by other interlocutors—for instance with direct agreement or disagreement, with conditionalization through propositional anaphora (as with "If so, then ...")', or with testimonial reports. Contents that are merely presupposed, implicated, entailed, or otherwise manifested can be targeted only by redirecting the conversational focus, by saying something like 'Hey, wait a minute! You seem to be assuming/suggesting that … '. Thus, establishing that the central point of ethical discourse is expressive rather than descriptive, as 'pure' expressivists aim to do, requires demonstrating that the at-issue moves proffered by canonical ethical utterances are non-descriptive. By contrast, hybrid expressivists have more flexibility here, since they may locate expressive commitments as either at-issue in combination with a descriptive claim, or as outside the conversational focus.

In the section "How Do Words Express?", I look more closely at the implications of the distinction between at-issue and peripheral meaning for the analysis of individual words. In the remainder of this section, I consider what expressing itself might be: what do speakers *do* when they express, as opposed to describe?

The simplest version of expressivism, both in terms of the attitude expressed and the mode of expression, is *emotivism*, which treats sentences like 'Murder is wrong' as "ejaculations," much like grimacing or saying 'Ugh!' or 'Boo!' (Ayer 1936, 103). As Stevenson (1937, 23) says, emotive meaning is "the tendency of a word, arising through the history of its usage, to produce (result from) affective responses in people. It is the immediate aura of feeling which hovers about a word." Ayer held that ethical statements are strictly meaningless, because he countenanced only descriptive, and specifically verificationist, linguistic meaning. But emotivism is problematic even independently of these highly controversial assumptions, because it treats the connection between psychological states and utterances in ultimately *causal* terms, much like the connection between smoke and fire. As Grice (1957) and Dretske (1981) noted, one state of affairs, $x$, may indicate another, $y$, in virtue of being reliably connected to it; in such cases, we may say that $x$

means *y*, and we may use *x* to draw inferences about *y*. But this is very different from the kind of meaning that words, or even non-conventional gestures like pantomime, have. A hallmark of 'non-natural' meaning, whether linguistic or mental, is that it can come apart from how the world is. Thus, one can say, or sign, that the house is on fire even if it isn't and even if one doesn't believe (or desire) that it is; by contrast, smoke can't be wrong about, or want, fire—it just *is*. Emotivism treats ethical statements as natural signs of emotional states. But people are all too capable of misrepresenting what they take to be right or wrong. More importantly, they frequently disagree with one another, treating each other's ethical commitments as wrong. Emotivism denies all this.

Emotivism does capture a key intuitive aspect of expressing: that it involves a kind of *showing* which is more direct than describing. But it goes too far in construing the relation between attitude and utterance in purely causal terms. A more flexible construal appeals to the idea of 'avowal'. Simon Blackburn (1998) articulates the idea thus:

> So what at last is said when we say that something is good or right? ... . We can now say ... what is done when we say such things. We avow a practical state. 'Avowal' here means that we express this state, make it public, or communicate it. We intend coordination with similar avowals or potential avowals from others, and this is the point of the communication. When this coordination is achieved, an intended direction is given to our joint practical lives and choices.
>
> (1998, 68–69)

Blackburn doesn't spell out what is involved in 'making public' here. One useful model is offered by Mitchell Green's account of self-expression as "showing how things are within" (2007, 106; see also Bar-On 2004). On Green's view, expressive behaviors are indeed *grounded* in or reliably caused by a certain inner state, but they need not be involuntary. More specifically, because they have the function (whether by evolution or intention) of showing those states, they thereby serve as *signals* of them to others. When an agent produces a signal of a state that they do have, that signal shows, and thereby expresses it; while if they produce the signal in the absence of the state, they merely purport to express it. Finally, while some expressive behaviors are natural, others, like sticking out one's tongue (which functions as an expression of contempt in the United States but as an expression of humility in Thailand), are conventional; when an expressive signal is conventional, its use *commits* the agent to having the correlative inner state.

Unlike ejaculation, an 'avowal' model like Green's has the flexibility to allow that agents can express not just feelings but also desires, preferences, and intentions, some of which may be highly abstract and structured. Given the complexity of ethical discourse, this is a good thing. The problem is that on this model, agents also plausibly express *beliefs* when they make sincere factual assertions, since assertions are conventional devices which are reliably if not universally caused by beliefs, which function to signal those beliefs, and which commit speakers to having them. Of course, the expressivist need not adopt Green's view of expression. But any account flexible enough to encompass the abstractness and diversity of ethical discourse, and to account for lies and disagreement, seems likely to deliver the same result.

A common response here is to grant that *all* sincere declarative utterances express mental states. Indeed, Alan Gibbard says, this should be uncontroversial: "That words express judgments will, of course, be accepted by almost everyone" (1990, 84). On this view, what

differentiates distinctively expressive utterances from descriptive ones is the *kind* of mental state they express. Schroeder (2008a) calls this the "parity thesis," and argues that it entails a substantive meta-semantic view: the Lockean doctrine of "Mentalism," on which language inherits meaning from thought, rather than the other way around (or independently or in interaction). Mentalism is at least somewhat controversial, but it is also endorsed by many theorists of meaning especially those of a broadly Gricean orientation who hold that particular utterances express mental states by getting their hearers to recognize those states in a certain self-reflexive way; conventional meanings are then explained derivatively, as a "standard procedure" (Grice 1989, 233) for speakers to mean or express such states.

The Parity Thesis is plausible, but it puts the expressivist in a delicate position because by ruling out the most obvious place to establish a contrast between expressive discourse and straightforward assertion. It thereby shifts much of the explanatory burden onto non-cognitivism, since it seems that any differences between the particular species of expression involved in factual assertion and in more narrowly 'expressive' expression will be inherited from the types of attitude that are expressed. It also highlights the need to provide positive evidence, not merely that agents often have and communicate non-cognitive attitudes about ethical topics, but that ethical language constitutes a "standard procedure" for manifesting them, as opposed to either reporting their existence or communicating them non-conventionally.

The most plausible way for a Parity-endorsing expressivist to augment the theoretical resources and range of evidence available to them is to look outward, to utterances' effects, rather than just inward, to the attitudes that produce them. As Blackburn says, an important, perhaps essential function of ethical avowals is to coordinate joint practical activities. For this reason, expressivists have typically supplemented or replaced the idea that ethical utterances express feelings with a dimension of practical engagement. (And indeed, many theorists hold that emotions themselves have an essentially motivational function.) Thus, Stevenson (1944) holds that sentences containing 'good' function both to declare the speaker's approval and to exhort others to approve. Hare (1952) goes further, analyzing moral statements as "universal prescriptions" that entail imperatives for action. More recently, Gibbard (2003) treats normative utterances as proposing plans for how to live.

The Mentalist has a straightforward, expressivist-friendly explanation of what it means for an utterance to have such exhortative, imperatival, or promissory force: in keeping with their general theory of meaning, on which utterances function to manifest attitudes, they analyze exhortations and imperatives as manifesting desires that the hearer does something, and promises as manifesting intentions to do something. *While this analysis is plausible at a psychological level, it prima facie* mischaracterizes the role that imperatives and promises play in discourse, which is to actually, directly, place the hearer or speaker under an obligation. It also doesn't yet provide the expressivist with evidence for a distinctively expressive conventional mechanism by which avowals implement the more generic relation of expression, since as far as the Mentalist story goes, the speaker could be expressing their desire or intention by simply reporting it.

Given these problems, some expressivists have complemented Mentalism with aspects of a more 'dynamic' theory of meaning, drawing on Stalnaker's (1970, 2002) notion of common ground and Lewis (1979) metaphor of the conversational score. These views are amenable to expressivism, insofar as they specify sentences' conventional meanings, not in terms of descriptive truth-conditions, but rather of conversational effects or "context

change potentials." So, just as the declaration "You are now married" alters the context directly, by making it the case that the couple *is* married, rather than indirectly, by describing them as married, so the permissive "You may now kiss your spouse" directly changes the context so that the addressee is permitted to kiss, rather than describing them as kissing-eligible. As long as these conventional conversational effects are ultimately *explained* in terms of expressed attitudes, so too that non-cognitive psychological states do the fundamental explanatory work of grounding linguistic meaning, the expressivist can supplement Mentalism with a 'dynamic', non-descriptivist specification of what those effects are.

The most direct way for an expressivist to incorporate conversational effects into their account would be to treat ethical statements as disguised imperatives and to appropriate the standard dynamic analysis of imperatives, which treats them as updating the addressee's 'To Do List' (Portner 2004). On this analysis, imperatives directly assign an action the status of to-be-done-by-addressee, rather than indirectly obliging the addressee to act by adding information about the speaker's desires, or about what actions are obligatory, to the common ground. For reasons we'll explore in the next section, a direct implementation of the dynamic analysis is unlikely to succeed for thin ethical terms like 'is wrong'. But expressivist analyses of deontic modals like 'must' and 'might', as functioning to update the common ground of plans (Gibbard 2003) or to alter preference orderings among possibilities (Silk 2015), are similar in spirit. Meanwhile, expressivists about epistemic and probability modals like 'might' and 'probably' have argued along structurally similar lines that those terms function to 'test' the context set for coherence, or to alter accessibility relations among information states, or to advise a certain credence distribution (Blackburn 1980; Yalcin 2007, 2012).

All of these analyses capture a way that certain classes of utterances might conventionally *do* something other than contribute information to the conversation, in a way that mirrors and is explained by having a non-doxastic mental attitude. They thereby give us a better grip on how one might demonstrate that a class of utterances has a non-descriptive function. At the same time, the mere fact that these utterances do have one non-descriptive conventional function doesn't itself establish that they don't also play a descriptive role. Further, the cases of epistemic and probability modalities mark dramatic departures from the simple emotivist model, with its intimate connections to feeling and practical action. Finally, they raise the question whether ethical predicates like 'good' and 'wrong' really do have analogous 'dynamic' effects.

I turn to these challenges in the following section. Summarizing the discussion to this point: to establish their view, the expressivist needs to provide evidence that canonical utterances involving the target class of words have a conventional function of expressing an attitude other than belief. Pure expressivists, unlike hybrid expressivists, also need to show that the at-issue contribution of such utterances is *not* to express belief. For both, the best place to seek this evidence is not just upstream, to the cluster of beliefs, desires, feelings, and intentions that motivate speakers to produce those utterances, but downstream, to their conventional conversational effects.

## HOW DO WORDS EXPRESS?

So far, we've focused on expressing as a relation between agents and whole utterances. We now turn to how language, as a conventional compositional communicative system, might imple-

ment this expressive relation, and in particular, how particular words might be 'semantically fitted' to perform it. Gibbard (2003, 7) advises, plausibly enough, that "to explain the meaning of a term, explain what states of mind the term can be used to express." While this seems innocuous, there is an immediate problem: most words don't express states of mind at all—at best, they express concepts, which combine to determine any of an indefinitely wide range of mental states. How do we identify a distinctively expressive role for a *word*?

The obvious strategy is to build up from simple cases. As Gideon Rosen (1998) suggests in, explaining Blackburn's 'quasi-realist' expressivism,

> The centerpiece of any quasi-realist 'account' is what I shall call a psychologistic semantics for the region: a mapping from statements in the area to the mental states they 'express' when uttered sincerely. The procedure is broadly recursive. Begin with an account of the *basic states*: the attitudes expressed by the simplest statements involving the region's characteristic vocabulary. Then assign operations on attitudes to the various constructions for generating complex statements in such a way as to determine an 'expressive role' for each of the infinitely many statements in the area.
>
> (1998, 387–88)

So, if we can specify the attitudes that are expressed when individual words like 'wrong' and 'good' are predicated of noun phrases like 'murder' and 'saving babies', we can then define functions on those basic attitudes that systematically relate them to other, more complex attitudes, in ways that mirror the relations among atomic sentences and phrases complex like "If … then" in language. To accomplish the first step, of assigning attitude-potentials to basic terms, we follow the same procedure of "reverse engineering" as for any other word: first, we survey a broad range of the target word's actual uses to identify a canonical subclass where it is used literally and sincerely; second, we extrapolate a stable contribution which that specific word makes to utterances of sentences combining it with other words; and finally, we posit a constant meaning by which it makes that contribution.

We ended the last section with the idea that at least one feature distinguishing ethical thought and talk from descriptions is a kind of motivational or imperatival force. How might an expressivist capture this intuition semantically? Linguistically, imperatival force is expressed by a sentence's grammatical mood, as in 'Bring me an umbrella!' *Reductivist* analyses of mood treat imperatives as disguised declarative sentences, like 'I command you to bring an umbrella' (Lewis 1970), or 'The next sentence is imperatival in force: You will bring an umbrella' (Davidson 1979). Such views are obviously unsuitable for the expressivist—as well as empirically and theoretically inadequate in their own terms (Starr 2014)—because they eliminate force as a linguistic phenomenon, treating it either as a merely pragmatic performance or as part of a sentence's truth-conditions.

A more moderate view treats grammatical mood as an 'illocutionary force marker', denoting an operation that takes a complete proposition—for instance, *that you bring me an umbrella*—as input and delivers an illocutionary act—say, a command—as output (Frege 1893; Searle 1969). Obviously, normative words aren't themselves grammatical moods. But dynamic analyses of epistemic, probability, and deontic modals like 'must', 'might' and 'ought' adopt a path that is structurally analogous, insofar as they too take whole propositions—model as input and deliver "context changes" as outputs, for

instance of altering the range of what it is obligatory to do. However plausible this may be as an analysis of for modal terms, extending it to predicates like 'wrong' and 'good' "is not straightforward, given that such terms take noun phrases like 'murder', rather than whole sentential clauses, like 'You come to work on time', as inputs. For such predicates, a model like Portner's that analyzes imperatives as denoting properties rather than whole propositions might seem more amenable. However, it is not obvious how much this helps, because the distinctive dynamic effect of modifying the addressee's To Do list is achieved, not by the type of property denoted, but rather from the fact that its argument is syntactically restricted in such a way that only the addressee can make it true—a feature that is not shared by statements like 'Murder is wrong'.

If imperatives don't offer a suitable linguistic model for assigning expressivist values to individual predicates, perhaps we should turn to a class of terms that do, by themselves actually, express states of mind: loaded words like 'damn' and 'jerk'. Chris Potts (2005, 2007) offers the leading linguistic analysis of (what he calls) 'expressives', which implements a basically emotivist model using contemporary semantic machinery. The central feature of Potts' account, both formally and substantively, is a robust segregation of expressive and descriptive meaning. Formally, he treats expressives as conveying a certain degree of positive or negative affect toward a subject. This contribution is non-descriptive, both in being simple and unarticulated, and in never altering any aspect of the context except the "expressive setting," which it updates directly, regardless of the word's location in the sentence. Expressive words and phrases that do contain descriptive content, like 'the damn dog', simply pass this content on, untouched, to the compositional machinery that determines a sentence's at-issue content and hence, its truth value.

How might an expressivist apply Potts' analysis to terms like 'wrong' and 'good'? Most expressivists would want to replace Potts' pure affective feeling with a non-cognitive attitude that is more structured and/or more practically engaged. While this is possible, Potts' analysis of *what* expressives express goes hand-in-hand with *how* he takes them to express it: it makes sense to segregate pure, "ineffable" feelings from descriptive contents syntactically because the two are so different substantively. And it is relatively plausible both that terms like 'damn' are devoid of descriptive meaning and that they don't interact significantly with the larger truth-conditional machinery. However, the deep reason why expressivists have moved away from emotivism is precisely that ethical and descriptive language are *not* segregated in these ways.

To see the problem with a segregationist analysis of ethical discourse, it will help to consider a range of alternative cases. Unlike 'pure' expressives, slurs (e.g. 'chink') clearly do make a substantive descriptive contribution as well as advocate an attitude—something like derogation of the target group. At the same time, these two contributions are still relatively distinct. Competent speakers can easily specify in (more) neutral terms which group is targeted by a slur; and typically only the predication of group membership contributes to the core at-issue content. This is especially clear with non-declarative utterances like bets and orders, which are intuitively satisfied if and only if the conditions determined by group membership are met (Camp 2013). It is also evidenced by the fact that when slurs are embedded within larger 'commitment-canceling' constructions—for instance, negations, questions, conditionals, modals, and indirect reports—typically only the descriptive content is 'bound' by the operator. The result is that on the one hand, the speaker doesn't end up committed to the actual application of that descriptive content, but instead to its negation, conditional

consequences, etc, while on the other hand the expressive element typically 'scopes out' so that the speaker *is* committed to the appropriateness of derogating the targeted group.

This separability of descriptive and expressive meaning has led a range of theorists to advance 'multi-dimensional' analyses of slurs that exclude their expressive meaning from the compositional determination of at-issue contentjust as Potts does for 'damn', treating that meaning variously as a conventional implicatures (Potts 2005, 2007; Whiting 2008; Williamson 2009), as conversationally implicated (Nunberg 2017), or as falling outside the realm of meaning altogether (Anderson and Lepore 2013). At the same time, though, slurs also differ from 'damn' in ways that challenge such a stringent segregation. Many theorists and ordinary speakers take assertions containing slurs to be false or incapable of truth, in contrast to sentences containing their neutral counterparts (Hom 2008; Richard 2008). And in an interesting range of cases, slurs' expressive element doesn't scope out of commitment-canceling constructions. Both of these features—effect on truth and binding within complex constructions—strongly suggest that with slurs, the expressive element *can* enter into the compositional determination of at-issue meaning, which is precisely what standard multi-dimensional model like Potts' is founded on denying (Camp 2017).

Other normatively laden terms are even more of a stretch for a robustly segregationist analysis. 'Thick' terms like 'lewd', 'fair', and 'courageous' are akin to slurs in being both descriptive and expressive. But unlike slurs, they lack lexicalized neutral counterparts; and in their case, the expressive dimension does appear to contribute to determining the term's extension—indeed, it is often claimed that descriptive and evaluative aspects are so intimately intertwined that one cannot grasp their truth-conditions without trying on the evaluative perspective (McDowell 1981; Williams 1985; Gibbard 1992).[1]

Finally, what about 'thin' ethical terms like 'good' and 'wrong'? A 'pure' expressivist, who holds that those words' meanings are exclusively non-descriptive, cannot follow Potts in completely separating expressive meaning from the compositional determination of descriptive, at-issue content since this would leave sentences containing those words devoid of any at-issue contribution at all. While Ayer embraced this conclusion, few contemporary expressivists do. Hybrid expressivists have more flexibility: they can posit a more or less minimal or schematic descriptive property as the compositional contribution to at-issue content, along with an expressive attitude that 'scopes out' of complex constructions. (For instance, the descriptive property might be one of meeting certain standards, and the expressive attitude might be one of endorsing those standards.) Indeed, some hybrid expressivists explicitly analogize thin ethical terms to slurs, adopting a conventional implicature view of both (Copp 2001, 2009; Boisvert 2008; see Schroeder 2009 and Teemu Toppinen's chapter "Hybrid Accounts of Ethical Thought and Talk", for discussion). Hybridism holds important explanatory advantages over pure expressivism. But, fundamentally, both pure and hybrid expressivists face a basic problem in deploying anything like Potts' formal model: thin ethical terms markedly *fail to* exhibit the remarkable independence from truth-assessment and from compositional involvement in complex constructions that is displayed by 'pure' expressives and, to a lesser extent, slurs.

To see the difference, consider 'If John is an *S*, I'll beat him to a pulp' and 'If John is an *R*, then you should hire him', where *S* is replaced by a pure expressive like 'bastard' and *R* by a slur like 'kike', and contrast both with 'If lying is wrong, then John would never lie'. In none of these cases is the speaker committed to the outright applicability of the predicate in the antecedent to John. Intuitively, though, a speaker of either of the first two sentences

*has* committed to the appropriateness of feeling negatively toward *S*s and *R*s; they've simply conditionalized on the applicability of feeling that way toward John in particular. By contrast, a speaker of the third sentence need not endorse any ethical standard at all: they might employ it in a general argument for amoralism, for instance. The same point goes for negation, questions, modals, and indirect reports: in all these cases, unlike with slurs and 'pure' expressives like 'damn', there is no incoherence in the speaker disavowing any particular set of feelings or commitments while maintaining their assertion of the complex sentence.

The lesson is a familiar one: thin ethical terms are fully and systematically entrenched within the compositional machinery of semantics. In essence, this is the Frege-Geach problem of specifying a consistent compositional contribution that an expressive word makes both to simple sentences and complex ones containing operators that block the speaker from commitments generated by a sincere utterance of the simple unembedded sentence (see Jack Woods' chapter "The Frege-Geach Problem", for discussion). This is precisely the problem that robust segregationism enables Potts to avoid. For 'pure' expressives like 'damn', and to a lesser extent for slurs, this seems appropriate. But the same does not hold for thin ethical terms.

The situation is not as hopeless as is sometimes made out. To accommodate the fact that stage actors and reporters don't actually make assertions, commands, etc., we already need to treat conventional markers of force as determining merely *potential* speech acts (Green 1997). Further, it sometimes is possible for force-contributing constructions, including imperatives, to embed, including in conditionals and disjunctions (Siegel 2006, Starr 2014). So we do need a theory of force-cancellation, including by embedding, anyway. Rather, the challenge is to specify what the relevant term *is* doing within such commitment-canceling contexts, and specifically to explain the appearance that it does much the same thing there as in simple sentences.

As with the interpretation of what it means to 'express' an attitude, one attractive strategy is to 'go global', by reinterpreting the existing logical machinery across the board in an expressivist-friendly way. The basic idea, pursued most systematically by Gibbard (2003) and Schroeder (2008b), is that attitudes like approving and intending are governed by a logic of consistency which is at least structurally analogous to the familiar truth-functional operations, and which ideally can be seen as a general schema of which truth-functions are just one instance. Making good on this project is already a significant challenge, one which Schroeder (2008b) argues is ultimately doomed to fail. But even if it succeeds, it is just one part of a larger story that must be told, given that: in addition to the logical operators, normative terms also occur smoothly in attitude and speech reports like 'X believes that P' and 'X asserts that P', in semantic claims like '"S expresses the proposition that P', and in alethic claims like 'It is true that P'. Thus, the meanings of all of these constructions, and the relations of entailment and inconsistency among them, also need to be explained in ways that don't essentially appeal to description and truth.

One way to accommodate the linguistic appearance of truth-conditionality while maintaining expressivism at a deeper level is to appeal to a deflationary or *minimalist* account of truth (Ayer 1936; Horwich 1990), on which asserting 'P is true' is equivalent to simply asserting P. The primary utility of the truth predicate, on this view, is to enable speakers to articulate generalizations about statements, as in 'Everything Janet said is true', rather than to ascribe a substantive property like correspondence to the world. Minimalism about truth is controversial; but if it works, it entitles the expressivist to adopt truth-talk without endorsing a substantively descriptivist psychology and ontology. Again, however, the problem is not limited to truth: the intimate interactions

among P, asserting P, believing P, and knowing P, and among asserting 'It is true that P', 'It is a fact that P', and so on, mean that the expressivist must also embrace minimalism about assertion, belief, knowledge, and reality, in what Dreier (2004) calls "the problem of creeping minimalism" (see also Rosen 1998). The global reach of the 'mission creep' involved in accommodating the role of normative words within more complex discourse thus threatens to render 'quasi-realists' like Gibbard (2003) and Blackburn (1993) and 'cognitivist expressivists' like Horgan and Timmons (2006) nearly indistinguishable from sophisticated, naturalistically oriented cognitivists and realists like Railton (1986) and Brink (1989). The expressivist (as well as the cognitivist and the realist) must then explain what differentiates expressivism from a 'robust' analysis of all these phenomena.

The most popular response, endorsed by Gibbard (2003) and Dreier (2004) among others, is "the explanation explanation," summed up in Blackburn's dictum that "In the philosophy of these things it is not what you say, but how you get to say it that determines your 'ism'" (1998, 7). Proponents of the explanation explanation grant that a semantics for thin ethical terms can be articulated in ('minimalist') descriptive terms, and that ethical statements function to express ('minimalist') beliefs, which can be true. So far, descriptive and normative statements are on a par. But expressivists argue that in the normative case, in contrast to the factual one, the best *meta-semantic* explanation of how sentences come to have these contents makes no appeal to moral facts, but appeals instead exclusively to psychological attitudes instead. By contrast, although the moral realist's explanation will presumably also cite normative beliefs, they will take these beliefs to be explained in turn by relations to moral facts—just as both parties agree that the explanation of straightforward factual claims ultimately appeals to facts about the 'natural' world. Further, the expressivist holds that this difference in the ultimate explanations of normative and factual statements arises *because of* differences in the functional roles of the attitudes they each express: the job of (robust, factual) beliefs is ultimately to track how the world is independently of the agent, while the job of desires, intentions, and other non-cognitive attitudes is to produce action.

Suppose that a global response to both the Frege-Geach problem and the problem of creeping minimalism along these lines succeeds, so that expressivism turns out to be both formally coherent and at least theoretically distinct from descriptivism. Now the opposite worry arises: that in its zeal to accommodate the apparently truth-involving contours of ordinary ethical talk, the global approach has eliminated the possibility of any direct positive evidence for expressivism. If the difference between descriptivism and expressivism turns entirely on different theoretical interpretations of the ultimate grounding of 'truth', 'belief', and 'fact', what remains of the core intuition that the two sorts of language work differently in ordinary discourse?

If the expressivist could demonstrate that non-cognitive attitudes themselves interact in ways that depart from straightforward descriptive beliefs, this would support the claim that they play a distinctive role in thought, which might then be reflected in talk as well. To this end, Charlow (2015) argues that the interactions we actually observe among those attitudes, especially with respect to undecided and uncertain states of mind, are such that they will be mischaracterized by any theory which identifies agents' attitudes by specifying their *contents*—even non-truth-conditional contents, like Gibbardian hyperplans—and then defines logical operations on those contents using Boolean relations of intersection, union, and exclusion (see also Schroeder 2008b; Silk 2015). However,

Charlow also argues that the standard logical operations are inadequate to ordinary factual talk as well, and concludes that we need a global reconstruction of semantics which replaces truth and consistency with broader categories of support and coherence. Thus, even if we do have reason to embrace a more expressivist-friendly semantics in general, it is not clear that this enables us to recover a strong contrast between descriptive and expressive language.

An alternative road to resuscitating a robustly contrastive expressivism challenges the assumed connection between ethical disagreement and assessment for truth. Thus, Khoo and Knobe (2016) provide experimental results suggesting that, in contrast to factual disagreement, people do not always take moral disagreement via negation to require one speaker's claim to be incorrect. If this is right, then the global assimilationist's appeal to minimalism may have prematurely written off a crucial source of evidence against standard descriptivist accounts—although it is less clear whether this supports expressivism in particular as opposed to a contextualist or relativist form of descriptivism.

## CONCLUSION

In considering how individual words might make an expressive contribution to larger sentences, we have arrived at much the same place as we did in investigating what it means to express a non-cognitive attitude. Expressivism seems like a coherent and attractive way to capture, within the analysis of language, many of the psychological and metaphysical intuitions that motivate non-cognitivists and anti-realists. The contrast with descriptive statements and beliefs is intuitive, and seems potentially explanatory. But the most plausible ways to articulate that contrast also threaten to undermine it, either by assimilating expression within the fold of description, beliefs, and truth, or by assimilating description, belief, and truth within the fold of expression and dynamic effects. Hybrid expressivists have more flexibility in explaining the intimate interactions between descriptive and expressive commitments than pure expressivists do, because they can locate non-cognitive attitudes outside the at-issue contributions of the compositional semantic machinery while retaining the claim that the target words conventionally function to express those attitudes. But their very flexibility also makes them harder to distinguish from cognitivists.

Expressivists of both stripes need positive evidence that expressive words have a conventional non-descriptive function, and a specification of how that function is implemented linguistically. Some of the most exciting recent developments within expressivism have focused on words and attitudes that are intimately connected with information and truth. Proponents of these views have brought important empirical data and formal resources from linguistics to bear; but the feasibility of marshaling these insights in support of classic expressivism about thin ethical terms is less obvious. At the same time, many of the most exciting developments within formal semantics concern aspects of language other than the at-issue presentation of descriptive content. Hybrid expressivists have begun to exploit some of these theoretical resources, but further interaction with linguistic theory is called for. From this perspective, expressivism about thin ethical terms may turn out to be the opening wedge of a more encompassing reconceptualization of semantic content, discourse structure, and communication.

## NOTE

1. See Väyrynen 2014 for arguments against a semantic analysis on the basis of projective behavior, and Kyle 2013 for defense of a semantic treatment.

## ACKNOWLEDGMENTS

Heartfelt thanks to David Plunkett, Tristram McPherson, and Lea Schroeder for extensive and extremely helpful comments. Thanks also to David Beaver, Dan Harris, Josh Knobe, and Eliot Michaelson for useful discussion.

## REFERENCES

Anderson, Luvell and Ernie Lepore (2013): "Slurring Words," *Noûs* 47:1, 25–48.
Austin, John (1961): "Performative Utterances," in *Philosophical Papers*, ed. J. O. Urmson and G. J. Warnock (Oxford: Clarendon Press), 233–252.
Ayer, Alfred Jules (1936): *Language, Truth and Logic* (New York: Dover).
Bar-On, Dorit (2004): *Speaking My Mind: Expression and Self-Knowledge* (Oxford: Clarendon Press).
Blackburn, Simon (1980): "Truth, Realism, and the Regulation of Theory," *Midwest Studies in Philosophy* 5:1, 353–372.
Blackburn, Simon (1993): *Essays in Quasi-Realism* (New York: Oxford University Press).
Blackburn, Simon (1998): *Ruling Passions* (Oxford, Clarendon Press).
Boisvert, Daniel (2008): "Expressive-Assertivism," *Pacific Philosophical Quarterly* 892, 169–203.
Brink, David (1989): *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press).
Camp, Elisabeth (2013): "Slurring Perspectives," *Analytic Philosophy* 54:3, 330–349.
Camp, Elisabeth (2017): "A Dual Act Analysis of Slurs," in *Bad Words*, ed. David Sosa (Oxford: OxfordUniversity Press).
Charlow, Nate (2015): "Prospects for an Expressivist Theory of Meaning," *Philosophers' Imprint* 15: 1–43.
Copp, David (2001): "Realist-Expressivism: A Neglected Option for Moral Realism," *Social Philosophy and Policy* 18: 1–43.
Copp, David (2009): "Realist-Expressivism and Conventional Implicature," *Oxford Studies in Metaethics* 4: 167–202.
Davidson, Donald (1979): "Moods and Performances," in *Meaning and Use*, ed. A. Margalit (Dordrecht: D. Reidel), 9–20.
Dreier, Jamie (2004): "Metaethics and the Problem of Creeping Minimalism," *Philosophical Perspectives* 18: 23–44.
Dretske, Fred (1981): *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press).
Frege, Gottlob (1893): *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet,* Vol. 1 (Jena: H. Pohle, 1st ed.)
Gibbard, Allan (1990): *Wise Choices, Apt Feelings* (Cambridge MA: Harvard University Press).
Gibbard, Allan (1992): "Morality and Thick Concepts," *Proceedings of the Aristotelian Society* Supplementary Volume 66, 267–283.
Gibbard, Allan (2003): *Thinking How to Live* (Cambridge, MA: Harvard University Press).
Green, Mitchell (1997): "On the Autonomy of Linguistic Meaning," *Mind* 106: 217–244.
Green, Mitchell (2007): *Self-Expression* (Oxford: Oxford University Press).
Grice, H. Paul (1989): *Studies in the Ways of Words* (Cambridge, MA: Harvard University Press).
Grice, H. Paul (1957): "Meaning," *Philosophical Review* 66:3, 377–388.
Hare, Richard Mervyn (1952): *The Language of Morals* (Oxford: Oxford University Press).
Hom, Christopher (2008): "The Semantics of Racial Epithets," *Journal of Philosophy* 105:8, 416–440.
Horgan, Terry and Mark Timmons (2006): "Morality without Moral Facts," in *Contemporary Debates in Moral Theory*, ed. James Dreier (Oxford: Blackwell), 220–238.
Horwich, Paul (1990): *Truth* (Oxford: Blackwell).

Khoo, Justin and Josh Knobe (2016): "Moral Disagreement and Moral Semantics," *Noûs*, online view; DOI:10.1111/nous.12151.
Kyle, Brent (2013): "How Are Thick Terms Evaluative?" *Philosophers' Imprint* 13:1, 1–20.
Lewis, David (1970): "General Semantics," *Synthese* 22:1/2, 18–67.
Lewis, David (1979): "Scorekeeping in a Language Game," *Journal of Philosophical Logic* 8:3, 339–359.
McDowell, John (1981): "Non-Cognitivism and Rule-Following," in *Wittgenstein: To Follow a Rule*, ed. S. Holtzman and C. Leich (London: Routledge and Kegan Paul), 141–172.
Nunberg, Geoff (2017): "The Social Life of Slurs," in *New Work on Speech Acts*, ed. Daniel Fogal, Daniel Harris, and Matt Moss (Oxford: Oxford University Press).
Portner, Paul (2004): "The Semantics of Imperatives within a Theory of Clause Types," in *Proceedings of Semantics and Linguistic Theory* 14, ed. Kazuha Watanabe and Robert B. Young (Ithaca, NY: CLC Publications).
Potts, Chris (2005): *The Logic of Conventional Implicatures* (Oxford: Oxford University Press).
Potts, Chris (2007): "The Expressive Dimension," *Theoretical Linguistics* 33:2, 165–197.
Railton, Peter (1986): "Moral Realism," *Philosophical Review* 95:2, 163–207.
Richard, Mark (2008): *When Truth Gives Out* (New York: Oxford University Press).
Ridge, Michael (2006): "Ecumenical Expressivism: Finessing Frege," *Ethics* 116:2, 302–336.
Ridge, Michael (2014): *Impassioned Belief* (New York, NY: Oxford University Press).
Rosen, Gideon (1998): "Blackburn's *Essays in Quasi-Realism*," *Noûs* 32:3, 386–405.
Schroeder, Mark (2008a): "Expression for Expressivists," *Philosophy and Phenomenological Research*, 76:1, 86–116.
Schroeder, Mark (2008b) *Being For: Evaluating the Semantic Program of Expressivism* (Oxford: Oxford University Press).
Schroeder, Mark (2009): "Hybrid Expressivism: Virtues and Vices," *Ethics* 119:2, 257–309.
Searle, John (1969): *Speech Acts* (Cambridge: Cambridge University Press).
Siegel, Muffy (2006): "Biscuit Conditionals: Quantification over Potential Literal Acts," *Linguistics and Philosophy* 29: 167–203.
Silk, Alex (2015): "How to Be an Ethical Expressivist," *Philosophy and Phenomenological Research* 91:1, 47–81.
Stalnaker, Robert (1970): "Pragmatics," *Synthese* 22:1/2, 272–289.
Stalnaker, Robert (2002): "Common Ground," *Linguistics and Philosophy* 25:5/6, 701–721.
Starr, Will (2014): "Mood, Force and Truth," *Language and Value. Protosociology* 31: 160–181.
Stevenson, Charles (1937): "The Emotive Meaning of Ethical Terms," *Mind* 46: 14–31.
Stevenson, Charles (1944): *Ethics and Language* (New Haven and London: Yale University Press).
Väyrynen, Pekka (2013): *The Lewd, the Rude and the Nasty: A Study of Thick Concepts in Ethics* (New York: Oxford University Press.
Whiting, Daniel (2008): "Conservatives and Racists: Inferential Role Semantics and Pejoratives," *Philosophia* 36: 375–388.
Williams, Bernard (1985): *Ethics and the Limits of Philosophy* (London: Fontana).
Williamson, Timothy (2009): "Reference, Inference, and the Semantics of Pejoratives," in *The Philosophy of David Kaplan*, ed. Joseph Almog and Paolo Leonardi (Oxford: Oxford University Press).
Yalcin, Seth (2007): "Epistemic Modals," *Mind* 116:464, 983–1026.
Yalcin, Seth (2012): "Bayesian Expressivism," *Proceedings of the Aristotelian Society* 112:2, 123–160.