Philos Stud (2022) 179:1–20
https://doi.org/10.1007/s11098-021-01648-7

Interpretative expressivism: A theory of normative belief

James L. D. Brown[1]

Accepted: 21 April 2021 / Published online: 12 May 2021

Abstract Metaethical expressivism is typically characterised as the view that normative statements express desire-like attitudes instead of beliefs. However, in this paper I argue that expressivists should claim that normative statements express beliefs in normative propositions, and not merely in some deflationary sense but in a theoretically robust sense explicated by a theory of propositional attitudes. I first argue that this can be achieved by combining an interpretationist understanding of belief with a nonfactualist view of normative belief content. This results in a view I call 'interpretative expressivism'. I then argue that traditional arguments employed by expressivists that normative statements express noncognitive attitudes can just as well support the claim that normative statements express nonfactual or nonrepresentational beliefs. Finally, I argue that this view has a number of advantages to versions of expressivism that deny that normative statements express non-deflationary normative beliefs.

Keywords Metaethics · Expressivism · Interpretationism · Belief · Nonfactual content

1 Introduction

Expressivism is the view that normative statements express nonrepresentational, nondescriptive, or nonfactual states of mind. Early versions of expressivism took this to mean that normative statements express emotions or desires (Ayer, 1936; Stevenson, 1937). More contemporary versions maintain that normative statements express sui generis normative attitudes with a distinctive nonrepresentational

✉ James L. D. Brown
jldbrown@hotmail.com

[1] University of Edinburgh, Edinburgh, UK

functional profile (Blackburn, 1998; Gibbard, 1990, 2003). Other versions claim that they express complex attitudes such as a belief-desire pair (Ridge, 2014). Although contemporary expressivists claim that normative attitudes can be beliefs in a minimal sense, what unites these views is the claim that, fundamentally, normative attitudes have a different psychological profile to prosaically factual beliefs.

This paper proposes a different approach. I argue that expressivists can and should maintain that normative statements express normative beliefs in a theoretically robust sense, where this is explicated by an independently plausible theory of propositional attitudes. This might sound to some like a non-starter. For it is natural to think that any robust conception of belief must be representational. However, although this assumption is widely shared by both expressivists and their opponents, it has received little scrutiny. I argue that the assumption can be questioned, and that rejecting the assumption opens the way for a suitably nonrepresentational theory of belief.

Specifically, I show how this can be achieved by combining a familiar expressivist strategy for explaining normative belief content with an interpretationist approach to explaining belief (Sect. 2). I then argue that the resulting view is better placed to solve or else avoid a number of problems faced by versions of expressivism that deny that normative and factual beliefs are fundamentally the same kind of attitude (Sect. 3). The interpretative expressivist picture not only improves upon existing versions of expressivism, however. If correct, it also challenges widely held assumptions about the representational nature of belief as such. The resulting view should therefore be of interest not only to metaethics, but to the philosophy of the mind and intentionality more generally.

2 The view

In this section, I explain how expressivism about normative thought can combine with an interpretationist theory of propositional attitudes to provide a suitably nonrepresentational theory of belief. First, I set out the interpretationist framework. Second, I introduce the notion of nonfactual content for normative attitudes. Third, I show how nonfactual contents can be employed within the interpretationist framework to provide an expressivist theory of normative belief.

2.1 Interpretationism

Interpretation is the process of ascribing beliefs, desires, and other propositional attitudes to a subject on the basis of what they do and say. Suppose Meredith is running across Covent Garden. Why is she doing this? Because she *wants* to see *La Traviata* and *believes* it's about to start. She hasn't picked up her ticket yet, so we can predict that she'll first go to the box office as she *knows* she won't get in without a ticket. Unfortunately, Meredith has forgotten her glasses and so she *hopes* that her seat isn't too far back. Upon *discovering* that she is up near the gods, we can expect Meredith to be *disappointed*. However, music lover that she is, we also expect her to overcome her disappointment and tell all her colleagues the day after that it was a wonderful performance. The predictions and explanations offered by such a narrative are not infallible and require a background of other assumptions. But the idea that we ascribe attitudes to others (and ourselves) in the service of the explanation and prediction of action should be familiar enough.

Interpretationism holds that reflection on the nature of interpretation can shed light on the nature of the mental. The label covers a fairly diverse family of theories (most prominently Davidson, 1980; Lewis, 1984; Dennett, 1987a; see Child, 1996 for an overview). Here, I will adopt something close to Dennett's *intentional stance strategy*. Dennett (1971) proposes that we make sense of this kind of interpretation by distinguishing between three different predictive strategies. First, we have the *physical stance*, which involves predicting behaviour on the basis of the physical state of an object together with the laws of nature. Second, we have the *design stance*, which involves predicting behaviour on the basis of an object's function (e.g. artefacts and biological objects). The design stance allows us to make predictions without any knowledge of the physical underpinnings of the object in question. For example, I do not need to know the physical laws governing my alarm clock in order to predict when it will ring. Third, we have the *intentional stance*, which involves ascribing beliefs and desires (i.e. intentional states) to an object on the assumption that it is a rational agent, and then predicting it to behave in ways that are rational given its beliefs and desires. This is the stance we took towards Meredith and that we utilise in folk psychology more generally. Although less reliable than the other stances, the intentional stance allows us to effortlessly understand and predict a vast amount of human behaviour, a task of otherwise immense complexity.

The intentional stance is first and foremost an epistemology of propositional attitudes. Dennett's additional metaphysical claim is that for an agent to possess propositional attitudes *just is* for that agent to be "reliably and voluminously" predictable using the intentional stance (1987b: 15). Thus, all it is for an agent *A* to believe that *p* is for the best (i.e. most predictive) interpretation of *A* to assign to *A* the belief that *p*. Importantly, however, whether an agent believes some proposition is an objective matter, as the predictive success of the interpretation is grounded in real patterns of behaviour (Dennett, 1987b: 25ff; 1991). While interpretationists like Dennett typically accept that there will also be a certain amount of indeterminacy about what an agent believes, this is seen as reflective of the phenomena. Overall, Dennett's interpretationism can be seen as a kind of dispositional view of belief: for an agent to believe some proposition is for that agent to be disposed to behave in certain ways under certain conditions (Dennett, 1987c: 50). Spelling out exactly what ways and under what conditions is the task of *intentional systems theory*, or *the theory of interpretation*, which aims to make explicit the rules of attribution implicit in folk psychological practice.

An initial gloss on these rules can be given as follows (Dennett, 1987b: 17):

> Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you

> predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many—but not all—instances yield a decision about what the agent ought to do; that is what you predict the agent *will* do.

Thus according to Dennett, the interpretative principles that constitutively govern propositional attitudes are *rational* principles about what agents ought to think, want, and do.[1] Beliefs and desires are therefore constitutively rational in that they are dispositions whose manifestation conditions are specified in terms of norms of rationality. Accordingly, we can understand what it is for something to be a belief, say, rather than a desire, in terms of the respective norms by which it is constitutively governed.

What exactly are these norms? While Dennett takes the assumption of rationality to play a "crucial role" (1987d: 96) in his theory, he claims that the relevant concept of rationality is "systematically pre-theoretical" and broad in scope (1987d: 98; Davidson, 1980: 241 and Lewis, 1986: 38f, 1994: 320ff make similar claims). Because of this, Dennett resists any attempt to provide a precise characterisation of the relevant notion of rationality. Nonetheless, he provides a number of general suggestions as to the kind of norms constitutive of rationality. Most centrally, beliefs and desires are governed by the principle of instrumental rationality: rational agents will act in ways that satisfy their desires in light of their beliefs. Beliefs are also distinctively governed by norms of theoretical rationality, such as norms relating belief to truth and evidence. Thus, rational agents will have mostly true beliefs about those parts of the world they have had exposure to, relative to their perceptual capacities and interests (Dennett, 1987b: 19). Sophisticated agents will also acquire new beliefs via reasoning, and eliminate inconsistencies when brought to light, relative to interests and resource limits (Dennett, 1987b: 20). What desires a rational agent ought to have will be determined in part by the nature of the agent in question. In virtue of the biologically evolved nature of human beings, Dennett proposes that we attribute to people a stock of basic desires such as "survival, absence of pain, food, comfort, procreation, entertainment" and so on (Dennett, 1987b: 20; see also Lewis, 1994: 320). Derived desires can be attributed on the basis of what the agent believes, such as desiring what an agent takes to be a means to an already desired end.

And so on. I freely admit that there is a lot of room for debate about these details and the overall viability of interpretationism in the philosophy of mind more generally (for discussion see Dennett, 2009). But the general idea that I want to work with in addressing the specific topic of this paper is that we assign to an agent the propositional attitudes that best rationalise that agent's behaviour in light of their nature, capacities, and sensory history. *A* believes *p* just in case *A*'s dispositional

[1] Does it follow that attitude ascriptions are normative? My preferred view is that the 'ought' of constitutive rationality is descriptive rather than normative, though this is not to deny that 'rational' has normative uses—see Ridge (2014: ch.8). If attitude ascriptions are normative, then expressivists might need an expressivist meta-theory to explain the use of normative notions in their theory of normative attitudes (e.g. Gibbard 2012), though Blackburn (2002: 164ff) denies that the normativity of attitude ascriptions has this implication (see also Chrisman 2016: 204ff).

profile is best rationalised by *A* believing *p*, where this is fixed by the norms implicit in folk psychology. It is a matter of contention precisely what these norms are, and indeed if the relevant notion of rationality is ultimately codifiable (see Child, 1996: ch.2). While below I will make substantive claims about the distinctive norms that govern normative attitudes, their precise formulation and codifiability will not be at issue. Dennett's interpretationism and the theory of normative belief developed below are compatible with a range of answers to these questions.

Importantly, on the interpretationist view, insofar as one represents the world in believing *p*, one does so *implicitly* in virtue of one's dispositional profile. This raises the possibility that the dispositions involved in believing a normative proposition do not involve an implicit representation of a normative environment. To make good on this possibility, I will argue that expressivists can make use of an off-the-shelf notion of *nonfactual belief content* to individuate the relevant disposition that constitutes believing a normative proposition. Because interpretationism is a *non-reductive* theory of intentionality in which belief contents figure in the explanantia of attitude ascriptions, this will be central to the proposed account. First, however, it will be helpful to examine what conception of belief content is best suited to the intentional stance strategy more generally.

2.2 Belief content

Although the intentional stance strategy as such is not committed to any particular conception of belief content, it is arguably best served by some kind of possible worlds conception of content (Dennett, 1987e). This can be characterised as follows (Yalcin, 2018a: 24):

> **The possible worlds model of content**. The content of a state of belief is representable as a set of possible worlds, intuitively the worlds "left open" by what is believed. Propositions are sets of possible worlds, and the propositions an agent believes are those true with respect to all of those worlds the state leaves open.

So, for instance, if I believe that Tibbles is on the mat, the content of this belief can be given as the set of possible worlds in which Tibbles (or some counterpart thereof) is on the (contextually specified) mat. This particular belief is understood as a property of my total belief state, which is primary. Specifically, I believe that Tibbles is on the mat when the content of my total belief state is a subset of the proposition that Tibbles is on the mat. Of course, the possible worlds model has its problems. Most pressingly, it apparently fails to discriminate believing necessarily equivalent propositions and apparently fails to make sense of deductive belief acquisition. However, I leave such controversies aside here. While there are various ways of adding complexity to the model to account for these difficulties (see, for example, Stalnaker, 1987; Lewis, 1986: 30ff; Yalcin, 2018a), for ease of exposition I will stick with the basic possible worlds model for ordinary beliefs.

The possible worlds model fits well with interpretationism for a number of reasons. First, taking an agent's total belief state to be primary facilitates the holistic nature of interpretation. Second, because possible worlds propositions are unstructured, agents are not required to possess language or entertain structured thoughts in order to have beliefs. This allows for the intentional stance to be taken towards non-human animals. Third, the possible worlds model can capture indeterminacies in belief embraced by interpretationism. One source of indeterminacy might be belief fragmentation (Stalnaker, 1987: 86f; Lewis, 1986: 30ff). Another might be how according to the possible worlds model, the relevant domain of alternative possibilities is defined relative to the discriminatory abilities of agents (Stalnaker, 1987: 58; Dennett, 1987e: 207). Consequently, indeterminacy might arise when an agent's beliefs discriminate more or less finely than the proposition under consideration. Fourth, Dennett (1987c: 48f) stresses how we are intentional agents in virtue of our evolved nature. This 'agent-relativity' of content on the possible worlds model allows us to capture this continuity and its development.

Appreciating the way in which belief contents are individuated relative to an agent's discriminatory abilities is central to understanding the sense in which (factual) beliefs are representational for the interpretationist. When I believe that Tibbles is on the mat, the set of worlds in which Tibbles is on the mat makes explicit what is represented by my belief state. However, in believing this proposition, I represent Tibbles being on the mat *implicitly* in virtue of my sensory and behavioural relation to my environment. It is in virtue of my behaviour and my sensory capacities that it is rational to attribute a certain discriminatory ability, viz. the ability to divide up a domain of possibilities into those in which Tibbles is on the mat and those in which Tibbles is not.

This is significant because it allows the expressivist to argue that the particular discriminative ability involved in believing a normative proposition consists in something other than an ability to discriminate between alternative factual possibilities. To illustrate this claim, we first need some nonfactual conception of normative content to work with. I propose to use Gibbard's (1990) notion of a *system of norms*. In part this is due to its familiarity within the metaethical literature, but also because it builds on the possible worlds model already outlined, and so should be equally well suited to interpretationism. However, the general approach to belief developed below is not committed to any particular way of modelling normative content, so long as it is suitably nonfactual (for other nonfactualist approaches to modelling normative content see Dreier, 1999; Gibbard, 2003; Yalcin, 2012, 2018b; Silk, 2013; Schroeder, 2013; Charlow, 2014).

Gibbard's suggestion is to model the content of one's normative attitudes in terms of the systems of norms one accepts. Just as a possible world w provides a complete specification of a way the world can be, a system of norms n provides a complete specification of what to do, think, or feel in any conceivable circumstance. As a first approximation, whereas the content of a factual belief is given by the set of worlds left open by that belief, the content of a normative belief is given by the set of systems of norms left open by that belief. The intuitive idea is that, relative to some situation, norms provide prescriptions about what to do, think, or feel, and to accept some normative content is to be disposed to behave in accordance with those prescriptions. In cases where the norms in question are less directly concerned with action, the relevant prescriptions might include assenting to the claim in normative discussion and exhibiting the right sort of affective attitudes towards those (including oneself) that act against what is prescribed.

We can then enrich the possible worlds model to arrive at the following:

> **The world-norm model of content**. The content of a state of belief is representable as a set of <*w*, *n*> pairs, intuitively the <*w*, *n*> pairs "left open" by what is believed. Propositions are sets of <*w*, *n*> pairs, and the propositions an agent believes are those that hold with respect to all of those <*w*, *n*> pairs the state leaves open.

From a formal perspective, the world-norm model provides a conservative extension of the possible worlds model. First, like possible worlds, systems of norms can be constructed out of possibilia (Yalcin, 2012: 147; Stalnaker, 2014: 130). If we think of norms as mappings from situations to permissible outcomes, a system of norms can be defined as a function from sets of worlds that realise possible situations to the sets that realise the permissible outcomes. Second, standard interpretations of logical notions carry directly over. Where ϕ and ψ are arbitrary sets of <*w*, *n*> pairs:

negation = ϕ'
conjunction = $\phi \cap \psi$
disjunction = $\phi \cup \psi$
inconsistency = $\phi_1,..., \phi_n$ are inconsistent iff $(\phi_1 \cap ... \cap \phi_n) = \varnothing$
entailment = $\phi_1,..., \phi_n$ entails ψ iff $(\phi_1 \cap ... \cap \phi_n) \subseteq \psi$

Accordingly, the model treats logical complexity in factual and normative thought in exactly the same way. Third, proponents of the possible worlds model already accept the need for additional parameters to encode non-world-characterising information, such as *de se* information about one's temporal and physical location in the world (Lewis, 1979). So the idea of belief contents as tuples of different kinds of information rather than sets of worlds is already familiar.

Intuitively, by characterising logical space in terms of <*w*, *n*> pairs, belief contents can distinguish not only between ways the world can be, but also between what to do, think, or feel. Moreover, because the logic of <*w*, *n*> pairs is identical to the logic of sets of worlds, we have a straightforward explanation of logically complex normative content. Our task now is to show how we can put this model to work within an interpretationist account of propositional attitudes to defend the claim that normative commitments are nonrepresentational beliefs.

2.3 Interpretative expressivism

To recap, the view being offered is that beliefs are behavioural dispositions individuated by constitutive principles of rationality. What makes some attitude a belief and not another attitude is the particular principles by which it is constitutively governed and which supply the basis for interpretation. What makes it the case than an agent believes *p* rather than *q* is that her actions are best predicted by interpreting her as believing *p* rather than *q*. As such, beliefs are not intrinsically representational. Rather, that one represents the world in believing *p* is a corollary of the predictions this interpretation affords. In the previous section, I examined a view according to which belief contents encode not only factual information about an agent's environment but also nonfactual information of a directive nature. In this section, I show how this view can be incorporated within an interpretationist account of propositional attitudes to deliver an expressivist theory of normative belief.

The basic idea is this. Agents with normative attitudes require a richer interpretative framework than that provided by the possible worlds model of content. This fact is explained in terms of the rational principles that constitutively govern purely normative and purely factual beliefs. However, because both kinds of attitude are governed by the constitutive norms of belief *tout court*, both attitudes are fundamentally beliefs. This raises two questions. What are the constitutive norms of rationality that explain why normative belief is nonrepresentational? And why are these principles and not others the correct norms of rationality? I argue that by reinterpreting standard expressivist arguments about the nature of normative attitudes as arguments about the constitutive norms governing belief, the expressivist can answer both questions at once. My aim is not to defend these arguments but simply to provide expressivists with a strategy for defending the interpretative view. However, as these arguments are familiar ground for expressivists, this provides licence for optimism that expressivists can make good on this strategy.

The first claim to defend is that the interpretation of normative agents (i.e. agents with normative attitudes) requires (at least) the richer framework of the world-norm model of content. The justification for this claim must ultimately come from the constitutive principles of rationality governing our attitudes. I suggest two principles which taken together might provide this justification. The first principle comes from the debate about motivational internalism. Arguably, something like the following is true of first-person ought-beliefs (see Ridge, 2015):

> **Normative Internalism**. Necessarily, for any fully rational agent *A*, if *A* believes (of herself) that she ought to Φ in *C*, then *A* will intend to Φ in *C*.

Normative Internalism receives support from the observation that in a number of contexts, an agent's believing that she ought to Φ in *C* is sufficient to explain her Φ-ing in *C*. For example, if I come to believe that all things considered I ought to give more money to famine relief, and a UNICEF collector comes to my door to collect money, then one would expect me to give money (Smith, 1994: 6). If I did not give money, then we would search for some countervailing factor, such as an overriding desire to hold onto my money, or my not having any money to give. But if I do give money, this seems perfectly intelligible even if I have no particular desire to give to the collector. However, if I had no particular desire to give to the collector, then my giving money requires some other explanation. If Normative Internalism is true, then my action can be explained by my judgment (and the intention it brings with it). For the purposes of our discussion, we can suppose that Normative Internalism purports to articulate a norm of rationality that constitutively governs interpretation.

Some care needs to be taken to explain what is meant by the claim that *A*'s believing that she ought to Φ in *C* is *sufficient* to explain her Φ-ing in *C*. For it cannot mean that an interpretation that assigned *only* this belief could explain *A*'s action. This is because interpretation is essentially holistic. Whether an agent possesses some attitude always depends on the totality of her beliefs *and* desires. Moreover, it is clear that in the UNICEF example we implicitly attribute many other beliefs and desires in order to make sense of my action (e.g. the belief that the collector is from UNICEF, the desire that if I give the collector money then this will contribute towards famine relief, and so on). So the notion of one's normative belief "sufficiently explaining" one's action needs some other explication. The suggestion being offered is that an explanation of an action typically requires not merely the presence of some related desires but an independent desire *to do that thing* (under some suitable description). Normative Internalism then claims that if an agent believes that she ought to Φ in *C*, then, other things being equal, she will Φ in *C* regardless of whether she has a desire to do that thing.

Normative Internalism alone does not support expressivism because it leaves open the possibility that normative beliefs are both intrinsically motivating and representational. So we need a second principle to complete step one of the argument. Intuitively, the relevant principle needs to capture the Humean idea that 'cold representations' of the world just aren't the kind of thing that alone could rationalise action. If an agent's total belief state is given by the set of <*w, n*> pairs left open by what is believed, we can pick out an agent's factual beliefs as the subset of their beliefs with *purely factual* or *norm-invariant* content, where the *n* parameter is left idle and no *n* is ruled out.

The suggestion is then that these beliefs are constitutively governed by the following principle:

> **Representational Inertness.** Necessarily, for any fully rational agent *A*, action Φ and set of purely factual propositions P, *A*'s believing P is not sufficient to explain *A*'s Φ-ing.

Arguably, Representational Inertness captures what motivates the Humean view of belief once we drop the assumption that all beliefs represent reality. Importantly, the principle is compatible with the view that *some* beliefs are sufficient to explain actions. It is just that such beliefs cannot be purely factual. Thus, taken together, Representational Inertness and Normative Internalism support the claim that an agent's normative beliefs must be something other than belief in purely factual propositions, as they would be for any descriptivist metaethical theory.

If we individuate beliefs using <*w*, *n*> contents, the link between an agent's normative beliefs and her actions is reflected in the prescriptions of the set of systems of norms she accepts, which makes explicit what she accepts in virtue of her dispositional profile. While this provides us with a suitably nonfactualist view of normative belief, it's important to see that this is not entailed by the world-norm model itself. After all, the model is consistent with a realist view according to which *w* denotes reality stripped of its normative features and *n* denotes the norms prescribed by objective reasons (Sinnott-Armstrong, 1993: 300f; Kalderon, 2007: 73), as well as certain relativist views (e.g. MacFarlane, 2014). Rather, that normative belief is nonfactual follows from the fact that the disposition individuated by world-norm contents on this view contains no implicit representation of the agent's environment, i.e. the world she inhabits. In other words, the psychological profile of believing a normative proposition entails no ability to discriminate between normative ways the world might be and entails no sensitivity to normative facts or reality. Moreover, if we did take this sort of directive disposition to contain an implicit representation of the agent's environment, it would violate Representational Inertness.

What reason is there to accept Representational Inertness? One kind of consideration invoked by expressivists derives from Moore's open question argument. For example, Blackburn claims that for any factual belief we might have, "there would still be issues of what importance to give it, what to do, and all the rest. For we have no conception of a 'truth condition' or fact of which mere apprehension by itself determines practical issues. For any fact, there is a question of what to do about it." (Blackburn, 1998: 70) One might fairly wonder whether this is just a restatement of Representational Inertness rather than an argument for it. So the expressivist needs to find some non-question begging way to establish the thesis. That said, if the principle is a constitutive norm of belief, one might wonder what sort of argument would establish this. Compare: what sort of argument would establish that the principle of instrumental rationality was a constitutive norm of belief? One might think that the dialectical burden here is negative in the sense that the defender of Representational Inertness would need to show that putative explanations of an agent's actions in terms of purely factual beliefs alone are in fact irrational or unintelligible. Specifically, they need to argue that it is always rational or intelligible for an agent to accept some purely factual proposition and be completely unmoved by it (absent related desires). Exactly how this argument will go might depend on the nature of the putative normative facts so believed. And while it would be beyond the scope of this paper to survey these arguments, it seems to me that this is exactly the kind of argument expressivists are already in the business of providing. For instance, compare Horgan and Timmons (1992) on whether believing natural facts can play a motivating role and Dreier (2015a, 2015b) on whether believing non-natural facts can play this role.

Expressivists might also appeal to facts about disagreement to explain why normative beliefs should be interpreted as nonrepresentational. The problem arises in cases of fundamental disagreement where two individuals have a different conception of the subject matter in question and systematically respond to different features of reality (Björnsson, 2017: 277). For example, consider Hare's (1952: 148) example of the Missionary and the Cannibals. In Hare's example, the Missionary learns that the Cannibals have a term similar to 'good' in that it is a general adjective of commendation. However, whereas the Missionary applies the term to people who are meek and gentle, the Cannibals apply it to people who collect a large number of scalps. In such a case, the two parties have a radically different conception of what is good. However, it seems plausible that the two parties still disagree about what is good. This seems to stand in contrast to purely factual disagreement. Here, when two parties have a radically different conception of the subject matter, we interpret their respective beliefs to be about different things and so not in disagreement. This already gives us reason to accept the world-norm model of content. But what explains this difference? Expressivists might argue that because factual beliefs represent reality, we should interpret agents' factual beliefs as being about that aspect of reality it is most intelligible to interpret them as responding to. Radical divergence in factual beliefs therefore mandates different ascriptions of content. By contrast, because the expressivist claims that believing a normative proposition does not involve responding to any normative part of reality, radical divergence in normative beliefs is intelligible and thus permitted (compare Field, 2018: 15).

Again, this is not the place to settle this debate. My aim is rather to show that arguments traditionally used to support the claim that normative commitments are noncognitive can be used to support the claim that they are nonrepresentational beliefs. Given that expressivists hold something like Normative Internalism and Representational Inertness anyway, proceeding this way should look attractive to expressivists. In the remainder of this section, I argue that despite their differences both kinds of attitude are fundamentally beliefs because they are both governed by the constitutive norms of belief tout court, and so are fundamentally beliefs according to interpretationism.

First, both normative and factual beliefs are governed by the same norms of theoretical rationality that apply exclusively to beliefs. For example, the theory of interpretation involves the assumption that agents more or less follow certain rules of logic and reasoning (Dennett, 1971: 95). These rules apply to an agent's beliefs. Thus, we predict that an agent will acquire new beliefs and eschew old beliefs through inference and reasoning, as well as eliminating inconsistencies in light of new evidence. Given that sets of $<w, n>$ pairs follow the same logic as sets of worlds, all the same predictions can be made about how an agent will reason to and from beliefs with normative content. The non-reductive nature of interpretationism is important here, because there is no requirement to explain intentional notions like inconsistency in belief in more basic, non-intentional terms (for example, in terms of representational failure—we'll return to this below). Finally, both normative and factual beliefs are plausibly governed by some kind of truth norm such that beliefs constitutively aim at the truth, where this simply amounts to the claim that agents should believe that *p* only if *p*. Given this deflationary gloss, expressivists can and should accept this as a norm of normative belief (Sinclair, 2006: 256).

Further, the holistic nature of belief explains how an agent can acquire normative beliefs in virtue of responding to features of reality without the need for explicit reasoning. For example, if belief acquisition is something that applies to total belief states, then an agent who (a) accepts norms that rule out inflicting gratuitous pain and (b) perceives someone inflicting gratuitous pain can be predicted to (c) acquire the belief that person's actions are impermissible without the need for explicit reasoning. So the rules of interpretation pertaining to theoretical rationality apply just as much to beliefs with world-norm contents as beliefs with possible worlds contents. Given that normative agents are in fact disposed to act in these ways, this supports the claim that these attitudes are beliefs.

Moving to the constitutive norms of practical rationality that govern beliefs, consider the principle of instrumental rationality: that rational agents act in ways that would fulfil their desires given their beliefs. If ever there was a constitutive norm of belief, presumably this is it. It might seem less obvious that this applies to normative beliefs. First, if we accept Normative Internalism, then the paradigm case of normative beliefs motivating actions will be directly rather than via their interactions with desires or other attitudes. However, it might also be because the principle of instrumental rationality is often cashed out in metaphysically robust terms. For example, desire satisfaction is often cashed out in terms of the factual content of the desire being realised, where realisation is a metaphysically substantive notion. However, given the non-reductive ambitions of interpretationism, there is no requirement to reduce intentional notions to metaphysical ones. So we can still talk about an agent's desires being realised even if these desires involve nonfactual content.

For example, suppose Alex believes that tax avoidance is wrong. Other things being equal, it follows from Normative Internalism that we should predict that Alex will pay his taxes in full. However, suppose that things are not equal. Suppose that Alex has a strong desire to earn as much money as possible, and this desire leads him to avoid paying his taxes in full whenever the opportunity arises. After some soul searching, however, Alex comes to form an overwhelming desire to avoid wrongdoing. The next day it's time for Alex to complete his tax returns. What will he do? It would be reasonable to expect that Alex will pay his taxes in full, as we know he believes that doing so will be a way of 'bringing it about' that he avoids wrongdoing, where the content of his desire is specified in terms of the set of worlds in which he acts in ways required by the set of norms left open by first-order normative inquiry. So Alex's belief that tax avoidance is wrong is subject to the principle of instrumental rationality in much the same way as any of his purely factual beliefs are (compare Ridge, 2020: 3334; Beddor, 2020: 2795).

In sum, I have argued that our normative and factual beliefs are both subject to the same constitutive norms of belief tout court. It follows given intepretationism that both attitudes are fundamentally beliefs. I have also argued that additional constitutive norms governing normative and factual belief in particular explain how an agent's normative beliefs cannot be interpreted using factual contents alone. The resulting picture is a robust theory of belief according to which some but not all beliefs are representational. In contrast to expressivist theories that only allow for a deflationary or quasi-realist conception of normative beliefs, the view being offered here allows for a theory of normative belief in the full-blooded sense of the term that is compatible with expressivism. Further, if some beliefs are nonrepresentational, it follows that belief as such is nonrepresentational. While this might sound alarming to some, if we accept the interpretationist picture, I think this is less surprising than it might seem. If beliefs are fundamentally dispositions, then the sense in which our beliefs are representational is already derivative. And once we attend to the sorts of dispositions characteristic of normative belief, and we see that it is not obvious that such a disposition implicitly represents the world, we are further loosened from the grip of the representationalist picture.

As I have indicated at various points, there are many explanatory burdens that one would need to take on to fully defend the sort of view outlined here. But these are burdens that already exist for any expressivist theory, at least insofar as it deploys the same kind of arguments to support it. So if my arguments are correct, then the view outlined above is at least as good as expressivist theories that claim that normative and factual beliefs are fundamentally different kinds of attitudes. In the next section, I argue that a unified view is better than a bifurcated view.

First, however, I want to forestall a possible objection. Interpretationist theories are committed to some form of the principle of charity or humanity in order to avoid worries about radical indeterminacy. According to one prominent version of the principle, we must attribute mostly true beliefs to agents we interpret. However, it might seem that this stands in tension with expressivism, which has no such requirement for normative beliefs, especially given the emphasis on the possibility of radical disagreement. In response, it should first be noted that attributing widespread error in an agent's normative beliefs is compatible with attributing mostly true beliefs to that agent. Second, there are principled reasons for why we should attribute true beliefs for ordinary factual beliefs that do not carry over to the normative case. Specifically, we attribute mostly true factual beliefs because of what Davidson calls the Principle of Correspondence, which "prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances." (2001: 211) Because the expressivist denies that we respond to normative features of the world, there is no reason to expect them to have mostly true normative beliefs. And we can easily understand that if (say) we had had *their* upbringing, then we would also have come to have had their (false) beliefs. Third, a more plausible principle of charity links interpretation not simply with truth but intelligibility: "in ascribing beliefs, we should seek to optimize agreement between what S believes and what she ought rationally to believe, in light of her situation, her other attitudes, and the available evidence." (Child, 1996: 8) Thus spelling out the relevant requirements of rationality just is specifying the content of the principle of charity.

3 Pros

Most contemporary expressivists accept that normative statements express normative beliefs. This is typically embraced in the 'quasi-realist' spirit of accommodating the realist-sounding features of normative discourse within an expressivist framework (e.g. Blackburn, 1993). While such expressivists accept normative belief-talk 'at the end of the day', the challenge for their theory is to earn their right to normative belief-talk given a noncognitivist account of normative thought according to which normative judgments are fundamentally desire-like attitudes. This differs from the proposal set out above. The traditional approach is to start with the thesis that factual and normative beliefs are fundamentally different kinds of attitudes, and then go on to explain how both can be properly thought of under the title of 'belief'. My approach is to start with the thesis that factual and normative beliefs are fundamentally the same kind of attitude, and then go on to explain how this is compatible with an expressivist account of normative thought. Thus, the order of explanation is reversed.

The distinction between robust and deflationary or minimalist conceptions of belief is therefore not primarily about whether beliefs are representational. Rather, the distinction is related to explanation. Specifically, a robust conception provides an account of the nature of belief that explains various features of our beliefs. By contrast, a minimalist conception states what conditions must be met for some attitude to be classified as a minimal belief. But that it has these features is explained not by its being a minimal belief, but by its being some other kind of attitude.

Call any expressivist view that maintains that factual and normative beliefs are distinct types of attitude *bifurcated attitude expressivism* and any that maintains that they are of the same type *unified attitude expressivism*. Given that both approaches aim to accommodate normative belief, it is natural to ask which approach is comparatively more attractive. In this section, I argue that unified attitude expressivism has three important benefits: (i) it is more theoretically parsimonious; (ii) it is better placed to explain mixed attitudes; and (iii) it is better placed to explain inconsistency.

3.1 Simplicity

Unified attitude expressivism is simpler than bifurcated attitude expressivism in that the former postulates the existence of a single attitude type where the latter postulates the existence of two or more attitude types. Consider the beliefs that grass is green and that murder is wrong. According to unified attitude expressivism, these are two instances of a single attitude type that are distinguished by their contents. According to bifurcated attitude expressivism, these are distinct kinds of attitude. Whereas the former is a robust belief, the latter is some other kind of attitude.

The bifurcated approach as such is compatible with any number of ways of cashing out this distinction. Typically, however, contemporary expressivists explain the distinction in functional terms. Specifically, that normative beliefs have a distinctive practical role more akin to desires (e.g. Blackburn, 1998) or plans (e.g. Gibbard, 2003) than robust belief. For concreteness, I will assume some kind of functionalist account of the distinction between factual and normative beliefs. However, the arguments below apply mutatis mutandis to other ways of cashing out this distinction.

If normative and factual beliefs are fundamentally distinct kinds of attitude, it is likely that attitude types will multiply further. Indeed, certain versions of expressivism seem to imply an infinite hierarchy of such attitudes (Schroeder, 2008: 49ff). This will be (at least) to account for mixed attitudes, such as the belief that grass is green or murder is wrong. Given the normative component, this attitude cannot be fundamentally representational in nature. Given the representational component, it cannot be fundamentally practical in nature. So it seems that we need to introduce some third kind of attitude to account for mixed thoughts which is neither fundamentally representational or practical in nature.

While simplicity may not count for too much on its own, it is worth emphasising that bifurcated attitude expressivism posits differences where it appears that there are none. This is because we use all the same locutions to talk about normative beliefs as we do other kinds of beliefs. From a pre-theoretical perspective, normative beliefs seem to be just one among many other kinds of beliefs. Moreover, from a theoretical perspective, normative beliefs have all the same properties that are central to other kinds of beliefs, such as their logical and semantic properties, as well as a number of phenomenological properties (Horgan & Timmons, 2006). Given that we seem to have a unified explanandum, we should expect a unified explanans. But this is exactly what bifurcated attitude expressivism denies. Thus, simplicity here is not just a theoretical virtue in and of itself, but seemingly demanded by that which we are attempting to explain (though see Ridge, 2009 for an attempt to meet this challenge head on given a bifurcated view).

This point is strengthened when we observe that expressivists not only owe us an account of normative and mixed beliefs, but of other types of normative attitude, such as desires, hopes, doubts, presuppositions, etc. This is known as the many attitudes problem (Rosen, 1998: 393ff; Schroeder, 2010: 83f). If the expressivist explains normative belief as distinct from ordinary belief, then it would seem that she would have to provide piecemeal explanations of all other normative propositional attitudes as well. By contrast, unified attitude expressivism provides a unified account not only of belief but of propositional attitudes more generally, and so no such problem arises.

Indeed, a recent discussion of the many attitudes problem supports this conclusion. Beddor (2020) argues that because the functional role of desire can be defined in relation to belief, and because all expressivists have a working account of normative belief, normative desires can be straightforwardly defined in relation to normative beliefs. Further, because other propositional attitudes can be understood in terms of beliefs and desires, expressivists can construct other propositional attitudes out of these attitudes. However, while this solution is available to bifurcated views, they remain committed to the claim that normative beliefs and desires are distinct in kind to their representational counterparts. So the solution remains simpler for unified views, and bifurcated views still owe us an account of their similarity given their differences. Moreover, bifurcated views cannot simply assume at the outset that the attitudes are sufficient similar. This must be shown piecemeal for each attitude, which was just the original problem. So while bifurcated views might be able to solve the many attitudes problem in the way Beddor suggests (or some other way, e.g. Köhler, 2017), unified views avoid the problem altogether by positing only one kind of attitude type for factual and normative beliefs and desires.

3.2 Mixed attitudes

The mixed attitudes problem refers to the problem of explaining the nature of attitudes with both normative and factual content, such as the belief that stealing is wrong or it never causes pain. According to bifurcated attitude expressivism, to believe that stealing never causes pain is to be in a state with a representational functional role which aims to track the world, and to believe that stealing is wrong is to be in a state with a conative functional role which aims to guide action. If these are fundamentally distinct state-types, what type of state is the mixed belief?

A natural suggestion is that mixed beliefs are combinations of purely normative and purely factual beliefs. However, this suggestion fails to explain mixed disjunctions. In a nutshell, the problem is that because it is possible to believe a disjunction without believing either disjunct, a mixed disjunction cannot be adequately characterised by any combination of purely normative and purely factual beliefs. If it were, then it would imply that the agent always believed at least one of the disjuncts (Schroeder, 2015: 12ff; Charlow, 2015: 10ff; Starr, 2016: 373f). Perhaps the disjunctive state can be explained as some kind of inferential commitment (Blackburn, 1988; Chrisman, 2016: 178ff). However, we are now owed an account of this third state-type and how it relates to factual and normative beliefs.

By contrast, there is no special account needed for mixed beliefs if we adopt unified attitude expressivism. On the version I have outlined above, mixed beliefs are just beliefs with world-norm contents. The nature of such beliefs is explained in terms of the conditions under which the theory of interpretation says it is correct to attribute beliefs with this kind of content. And this will be explained fundamentally in terms of an agent's overall dispositional profile, where the attribution of a logically complex belief will involve dispositions to reason in certain ways. However exactly this explanation goes, it is not fundamentally different in kind to explanations involving purely factual or purely normative beliefs.

Thus, an agent *A* will believe a mixed disjunction *p* or *q* without believing *p* or believing *q* when the best interpretation of *A*'s actions assigns only the first belief and not the others. For instance: it is correct to not assign to *A* the belief *p* or the belief *q* if *A* is not disposed to act as if they believe *p* or believe *q*; and it is correct to assign the belief *p* or *q* if *A* is disposed to form the belief that *p* upon learning not-*q* and if *A* is disposed to form the belief that *q* upon learning not-*p*. Moreover, because believing *p* or *q* without believing *p* or believing *q just is* to be predictable in this sort of way according to interpretationism, there is no further question about what this belief must be like in order to have these properties.

It's worth stressing the importance of the holistic aspect of belief individuation here. Schroeder (2015: 21) argues that expressivists are not entitled to the world-norm model because they lack a sufficiently rich characterisation of the functional role of belief that can distinguish between arbitrary sets of <*w*, *n*> pairs. We can retain our grip on the functional role of normative beliefs by accepting the bifurcated approach, Schroeder thinks, but this just leads us to the disjunction problem. However, this objection presupposes that beliefs are individuated by their canonical or core functional role, understood as the subset of their total functional role that fixes their content. By contrast, the interpretationist view individuates beliefs relative to an agent's *total* belief (and desire) state, which is primary, where this latter assignment is governed by the totality of constitutive norms of rationality contained within the theory of interpretation. Thus, the richer and more powerful characterisation of belief comes not from the richer functional role of individual beliefs but from the holistic interpretative nexus by which we predict normative agents. Moreover, not only does the world-norm model allow us to make the predictions we need of normative agents, it makes their actions intelligible in terms of the distinctive rational principles that govern these predictions.

3.3 Inconsistency

Expressivists notoriously have a difficult time explaining how normative thoughts can stand in the right sort of inconsistency relations (Schroeder, 2008; Unwin, 2001; Wright, 1988). As we saw above, introducing nonfactual contents allows the expressivist to define two beliefs as inconsistent when the intersection of their contents is empty. But simply assigning informal contents with empty overlap to attitudes is not sufficient to show that those states are inconsistent in the right sense (Starr, 2016: 368f; Willer, 2017: 197ff). After all, it is not inconsistent in this sense to have desires with empty overlap. So we need to know *why* it is inconsistent to have beliefs with incompatible contents.

Assuming bifurcated attitude expressivism, one possible explanation would appeal to the core functional roles of each attitude type and explain how their constitutive functions are frustrated when an agent believes contents with empty overlap. In the factual case, we might say that beliefs with empty intersection necessarily fail to represent the world as being some way, because there is no way the world can be such that their contents are both true. Thus, factual inconsistency engenders a kind of representational failure. In the normative case, we might say that beliefs with empty intersection necessarily fail to determine a coherent prescription, because there is no way of acting such that both prescriptions can be realised. Thus, normative inconsistency frustrates the practical function of normative judgments to settle the thing to do.

There are a number of problems with this approach. First, one might worry whether the sort of practical inconsistency appealed to by the expressivist is of the right kind to ground *logical* inconsistency. If I plan to eat cake and to not eat cake, I presumably have an incoherent plan, but this does not obviously make me logically inconsistent. However, even when '*p*' denotes a normative proposition, it is surely logically inconsistent to believe both *p* and not-*p*. Second, even assuming a suitable sense of practical inconsistency is specified, we have a bifurcated explanation for a seemingly unified explanandum. This is because inconsistency in belief appears to be of the same kind regardless of whether the beliefs are normative or factual. Moreover, combinations of normative, factual, *and* mixed beliefs can be logically inconsistent. It is difficult to see how this could be true if normative and factual inconsistency were different in kind.

By contrast, because unified attitude expressivism counts all beliefs as instances of a single kind of attitude, inconsistency is explained as the same across normative, factual, and mixed contexts. Given our interpretationist theory of belief, how does this explanation go? The first thing to note is that given the theory's non-reductive commitments, we should not expect inconsistency in belief to reduce to some non-intentional notion, such as some kind of functional failure. Thus, trivially, to believe just is to believe as true, and believing inconsistent propositions necessarily results in believing a falsehood. If the theory of interpretation contains or entails some epistemic truth norm for believing only what is true, then inconsistent beliefs will violate this principle.

Further, consider that interpretation is based on how an agent acts, where the attitudes ascribed to the agent provide explanatory reasons that make her action intelligible. If beliefs are constitutively governed by rational principles, this suggests another way of thinking about what is wrong with believing inconsistent claims. Specifically, we ask what would be wrong with a fully rational agent who believes inconsistent claims. If we think of believing propositions in terms of ruling out sets of world-norm pairs, to believe inconsistent propositions is to rule out all doxastic possibilities. Which is to say the agent has no beliefs which could serve as reasons for action. Interpretation breaks down. Importantly, this explanation of what's wrong with believing inconsistent propositions applies equally to normative, factual, and mixed beliefs.

4 Conclusion

I have argued that expressivists can and should argue that normative statements express normative beliefs. I argued that interpretationism provides an attractive framework in which to give content to this claim. If successful, the interpretationist view provides a straightforward explanation of logically complex normative thought, mixed thoughts, and inconsistency—all of which expressivism traditionally struggles to explain. No doubt there is more to be said in each case, but these are nontrivial matters that face any theory of belief. Given that I have made use of an off-the-shelf notion of normative content, one might wonder why expressivists have not already considered this approach. A tentative diagnosis is that those engaged in metaethical debates have assumed that any theoretically robust notion of belief must be representational and that any other notion of belief must be deflationary and non-explanatory. I hope to have shown that it is open to expressivists to develop a suitably nonrepresentational theory of belief and other propositional attitudes.

Finally, I have not claimed that interpretationism is the only framework in which nonfactual contents can be put to work in developing an expressivist-friendly theory of belief. For example, 'common sense' or analytic functionalism (e.g. Lewis, 1994) shares many of the features of the interpretationist view endorsed here. Köhler (2017) provides a somewhat different functionalist account of belief for expressivists. Schroeder (2013) provides a unified non-cognitivist theory of belief as states of *being for*. And Gibbard (2012) can be read as providing a unified sentential theory of belief. So the interpretationist approach must ultimately be assessed in relation to rival approaches as well as on its own terms.

Acknowledgements Many thanks to Matthew Chrisman, Sebastian Köhler, James Laing, Brian Rabern, Michael Ridge, Debbie Roberts, Mark Schroeder, and an anonymous referee for helpful comments on earlier versions of this paper.

References

Ayer, A. J. (1936). *Language, truth, and logic*. London: Gollancz.

Beddor, B. (2020). A solution to the many attitudes problem. *Philosophical Studies, 177*(9), 2789–2813.

Björnsson, G. (2017). The significance of ethical disagreement for theories of ethical thought and talk. In T. McPherson & D. Plunkett (Eds.), *The Routledge handbook of metaethics*. London: Taylor and Francis.

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

Blackburn, S. (1993). *Essays in quasi-realism*. New York: Oxford University Press.

Blackburn, S. (1998). *Ruling passions*. Oxford: Clarendon Press.

Blackburn, S. (2002). Replies. *Philosophy and Phenomenological Research, 115*, 164–176

Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Charlow, N. (2014). The problem with the Frege-Geach Problem. *Philosophical Studies, 167*, 635–665

Charlow, N. (2015). Prospects for an expressivist theory of meaning. *Philosophers' Imprint, 15*, 1–43.

Child, W. (1996). *Causality, interpretation, and the mind*. Oxford: Clarendon.

Chrisman, M. (2016). *The meaning of 'Ought.'* New York: Oxford University Press.

Davidson, D. (1980). *Essays on actions and events*. New York: Oxford University Press.

Davidson, D. (2001). Three varieties of knowledge. In *Subjective, intersubjective, objective*. Oxford: Clarendon.

Dennett, D. (1971). Intentional systems. *The Journal of Philosophy, 68*, 87–106

Dennett, D. (1987a). *The intentional stance*. Cambridge, MA: MIT Press.

Dennett, D. (1987b) *'True Believers' in Dennett 1987a.*

Dennett, D. (1987c) *'Three Kinds of Intentional Psychology' in Dennett 1987a.*

Dennett, D. (1987d) *'Making Sense of Ourselves' in Dennett 1987a.*

Dennett, D. (1987e) *'Beyond Belief' in Dennett 1987a.*

Dennett, D. (1991). Real patterns. *The Journal of Philosophy, 88*, 27–51

Dennett, D. (2009). Intentional systems theory. In A. Beckermann, B. P. McLaughlin, & S. Walter (Eds.), *The Oxford handbook of philosophy of mind*. New York: Oxford University Press.

Dreier, J. (1999). Transforming expressivism. *Noûs, 33*, 558–572

Dreier, J. (2015a). Another world. In R. M. Johnson & M. Smith (Eds.), *Passions and projections*. New York: Oxford University Press.

Dreier, J. (2015b). Can reasons fundamentalism answer the normative question? In G. Björnsson, C. Strandberg, & R. F. Olinder (Eds.), *Motivational internalism*. New York: Oxford University Press.

Field, H. (2018). Epistemology from an evaluationist perspective. *Philosophers' Imprint, 18*, 1–23

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

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

Gibbard, A. (2012). *Meaning and normativity*. New York: Oxford University Press.

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

Horgan, T., & Timmons, M. (1992). Troubles for new wave moral semantics: The 'open question argument revived. *Philosophical Papers, 21*, 153–175

Horgan, T., & Timmons, M. (2006). Cognitivist expressivism. In T. Horgan & M. Timmons (Eds.), *Metaethics after moore*. Oxford: Clarendon.

Kalderon, M. (2007). *Moral fictionalism*. New York: Oxford University Press.

Köhler, S. (2017). Expressivism, belief, and all that. *The Journal of Philosophy, 114*, 189–207

Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review, 88*, 514–543

Lewis, D. (1984). Radical interpretation. *Synthese, 23*, 331–344

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell.

Lewis, D. (1994). *"Reduction of Mind". In Papers on metaphysics and epistemology*. Cambridge: Cambridge University Press.

MacFarlane, J. (2014). *Assessment sensitivity*. New York: Oxford University Press.

Ridge, M. (2009). Moral assertion for expressivists. *Philosophical Issues, 19*, 182–204

Ridge, M. (2014). *Impassioned belief.* New York: Oxford University Press.

Ridge, M. (2015). Evaluative judgments, judgments about reasons, and motivations. In G. Björnsson, C. Strandberg, & R. F. Olinder (Eds.), *Motivational internalism.* New York: Oxford University Press.

Ridge, M. (2020). Normative certitude for expressivists. *Synthese, 197*, 3325–3347

Rosen, G. (1998). Blackburn's *Essays in Quasi-Realism. Nous, 32*, 386–405

Schroeder, M. (2008). *Being for*. Oxford: Clarendon Press.

Schroeder, M. (2010). *Noncognitivism in ethics*. London: Routledge.

Schroeder, M. (2013). Two roles for propositions: Cause for divorce? *Noûs, 47*, 409–430

Schroeder, M. (2015). *Expressing our attitudes*. New York: Oxford University Press.

Silk, A. (2013). Truth-conditions and the meaning of ethical terms. In N. Shafer-Landau (Ed.), *Oxford studies in metaethics 8.*OUP.

Sinclair, N. (2006). The moral belief problem. *Ratio, 19*, 249–260

Sinnott-Armstrong, W. (1993). Some problems for Gibbard's norm-expressivism. *Philosophical Studies, 69*, 297–313

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

Stalnaker, R. (1987). *Inquiry*. Cambridge, MA: MIT Press.

Stalnaker, R. (2014). *Context*. New York: Oxford University Press.

Starr, W. B. (2016). Dynamic expressivism about deontic modality. In N. Charlow & M. Chrisman (Eds.), *Deontic modality.* New York: Oxford University Press.

Stevenson, C. L. (1937). The emotive meaning of ethical terms. *Mind, 46*, 14–31

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

Willer, M. (2017). Advice for noncognitivists. *Pacific Philosophical Quarterly, 98*, 174–207

Wright, C. (1988). Realism, antirealism, irrealism, quasi-realism. *Midwest Studies in Philosophy, 12*, 25–49

Yalcin, S. (2012). Bayesian expressivism. *Proceedings of the Aristotelian Society, 112*, 123–160

Yalcin, S. (2018). Expressivism by force. In D. Fogal, D. Harris, & M. Moss (Eds.), *New work on speech acts.* New York: Oxford University Press.

Yalcin, S. (2018). Belief as question-sensitive. *Philosophy and Phenomenological Research, 97*, 23–47