# Supervenience, Reliability, and the Explanatory Limitations of Conciliatory Expressivism

By Tristram McPherson (Ohio State University)
and David Plunkett (Dartmouth College)

Version of April 24, 2026
Forthcoming in *Philosophical Studies*
Please cite and quote final published version

## Introduction

Expressivism is an important position in contemporary metaethics. Roughly, it holds that the meaning of each ethical sentence is explained by the "desire-like" mental state (allegedly) conventionally expressed by use of that sentence.[1] Simple versions of expressivism fail to explain the apparent expressive power of ethical language. For example, A. J. Ayer's version of expressivism entails that ethical sentences cannot be true or false.[2] Many contemporary expressivists have taken such conflict to decisively favor what Sebastian Köhler has helpfully dubbed *conciliatory expressivism*.[3] Conciliatory expressivism uses resources such as "minimalist" interpretations of words including 'truth' or 'fact' to explain the apparent expressive power of ethical language.[4] Conciliatory expressivism is an important component of the "quasi-realist" expressivism endorsed by Simon Blackburn and Allan Gibbard, among others.[5] Put roughly, quasi-realist expressivism – which is arguably the leading form of expressivism in contemporary metaethics – aims to embrace several of the commitments of "realist" metaethical views, on an expressivist basis.[6]

Metaethicists have often claimed that expressivism either avoids or elegantly answers explanatory challenges about the metaphysics and epistemology of ethics in ways that "realist" metaethical views cannot.[7] However, in recent years, a number of philosophers – including Sharon Street, Jamie Dreier, Terence Cuneo, and Joshua Schechter – have cast doubt on the idea that conciliatory forms of expressivism have such an advantage with respect to certain prominent explanatory challenges.[8] In this

[1] For an overview of expressivism and its history in metaethics, see (Camp 2017) and (Schroeder 2010). Note that some use 'expressivism' more broadly, to encompass "non-descriptivist" views which do not explain meanings in terms of mental states.
[2] (Ayer 1936/1971).
[3] (Köhler 2021).
[4] Here and elsewhere in this paper, we use single quotes (such as 'truth') to mention linguistic items. We use double quotes (such as "minimalist") for a variety of purposes, such as quotation, mixes of use and mention, and "scare quotes".
[5] See (Blackburn 1993) and (Gibbard 2003). For an overview, see (Cuneo 2017).
[6] We return later in this paper (in §1.2) to discuss the relation between conciliatory expressivism and quasi-realism.
[7] Examples of challenges related to supervenience include (Hare 1952/1991), (Blackburn 1993), and (Gibbard 2003), examples of challenges related to reliability include (Blackburn Manuscript) and (Gibbard 2011).
[8] See (Street 2011), (Dreier 2015), (Cuneo 2008), (Cuneo 2020), and (Schechter Manuscript).

paper, we build on this work, focusing on two of the most prominent such challenges: the "supervenience challenge" and the "reliability challenge".[9] We discuss these challenges together because they share a deep structure: each is a challenge to explain modal correlations.

Our discussion does not attempt to exhaustively canvas every way that philosophers have attempted to implement conciliatory expressivism. Instead, it focuses on the relationship between the nature of these specific challenges and the general nature and ambitions of conciliatory expressivism.

We argue that there are powerful reasons to think that conciliatory expressivists have no distinctive resources that enable them to address these explanatory challenges. That is, to the extent that the conciliatory expressivists have such resources, it is solely in virtue of commitments that proponents of various competing views can share.

In outline, we argue as follows. First, there are just two fundamental ways of addressing the explanatory challenges we discuss: to *undercut* them (by showing that the relevant explanatory demand does not arise), or to *answer* them (by providing explanations of the relevant modal correlations). Second, the hypotheses about language that are necessary to achieve an even modestly conciliatory form of expressivism entail that the conciliatory expressivist cannot undercut these challenges in a straightforward way (§3). Third, the conciliatory expressivist's characteristic hypotheses about language and thought provide them with no distinctive additional resources that could be used to answer the challenges (§4). We conclude (in §5) with some general lessons for evaluating the prospects of expressivism and related views, both in metaethics and in other areas of philosophy where they have been advanced, such as epistemology, philosophy of law, and metaphysics. Before turning to this argument, however, we more carefully introduce the resources needed to prosecute it: our accounts of conciliatory expressivism (§1) and of the supervenience and reliability challenges (§2).

## 1. Expressivism and Conciliation

Our argument is based crucially on the defining commitments of conciliatory expressivism. These can be easy to misunderstand. It can also be easy to subtly conflate conciliatory expressivism with other views, which have quite different explanatory resources available to them. Because of this, we introduce the view more carefully, and more slowly, than is standard in this kind of context.

We understand conciliatory expressivism as the combination of two theses: 1) "pure" expressivism about ethical language and 2) a "conciliatory" account of certain other language. We discuss each in turn.

[9] For the supervenience challenge, see, e.g., (Blackburn 1993) and (McPherson 2012). For the reliability challenge, see e.g. (Street 2006) and (Enoch 2010).

### 1.1 Pure expressivism

Here, we focus on pure *metaethical* expressivism, which is a view about the character of our actual ethical thought and talk.[10] We understand pure expressivism (hereafter, for brevity, just "expressivism") as characterized by two key theses: non-cognitivism about ethical thought and mentalism about the meanings of sentences. We introduce these ideas in turn.

Non-cognitivism presupposes an influential, but controversial, distinction between "cognitive" and "non-cognitive" mental states.[11] This distinction is commonly spelled out in terms of the functional profiles of mental state types, characterized in terms of "direction of fit".[12]

On this view, *cognitive* states, such as beliefs, have a "representational" functional profile. This profile is characterized by a "mind-to-world" direction of fit: roughly, a cognitive state's functional teleology is to have a content that "fits" the relevant way the world is. For example, take Roberta's *belief* that there is water in the kitchen. Other things being equal, this state will tend to "fit the world" by being responsive to evidence concerning whether there is water in the kitchen.

By contrast, *non-cognitive* states, such as ordinary desires, have a "world-to-mind" direction of fit. Roughly, their functional teleology is to cause the agent to change the world to fit their content. For example, take Roberta's *desire to get some water*. Other things equal, this state will move her to get water.

With this distinction in hand, we can formulate non-cognitivism as the view that ethical thoughts, such *it is wrong to kill*, are non-cognitive states. That is, these states function not to represent the world, but rather to motivate behavior. Non-cognitivists disagree about the precise character of ethical thoughts, however. For example: some think that the above thought consists in something like *disapproval* of killing, while others take it to consist instead in something like a *plan* not to kill.[13]

Now turn to the expressivist's account of talk. As we understand it, expressivism presupposes a general view about the explanation of sentence meaning that we call *mentalism*. Put roughly, on this view, the meaning of a sentence is explained by facts about the mental state "expressed" by uttering that sentence. There is significant debate about how exactly to unpack what "expression" means here. For the purposes of this paper, we focus on the following prominent proposal:

**Expression** For utterance U in language L to express mental state M is for

---

[10] For our broader understanding of the metaethical project, see (McPherson and Plunkett 2017).
[11] For discussion, see (Bedke 2017). Here, as elsewhere, we grant expressivists the core assumptions needed for their view to be a live hypothesis.
[12] For more careful discussion, see (Faraci and McPherson 2017).
[13] Others deny that there is any particular attitude that one must have to have an ethical thought. Compare (Björnsson and McPherson 2014).

> M to be the condition on correct utterance of U that is given by the semantic norms in L.[14]

To illustrate mentalism, consider a simple sentence:

(S) 'Water is wet.'

Plausibly, to correctly assert (S) in English, one needs to (at least) have a certain *belief*: namely, the belief that *that water is wet*.[15] We will assume that this belief is the mental state *expressed*, in our technical sense, by uttering (S). In turn, this can be used to give a mentalist account of the meaning of S. This belief *represents* the world as being a certain way. We can think of the meaning of the sentence as the representational content of that belief. On a familiar view, this would be the proposition that water is wet.

Next consider:

(H) 'Hooray for Coco!'

Arguably, the semantic correctness condition on uttering (H) is something non-cognitive, such as *having a positive attitude towards Coco.* Unlike for (S), we cannot understand the meaning of (H) in terms of the content of the representational state expressed. This is because the expressed state is non-cognitive rather than representational. And this seems like a virtue of the mentalist picture: for (H) appears meaningful, but not to purport to tell us how the world is.

Expressivism combines mentalism about the meaning of speech acts with non-cognitivism about ethical thought. To illustrate this combination, consider an ethical sentence:

(E) 'It is wrong to kill.'

According to pure expressivism, the semantic correctness condition for (E) involves being in a certain non-cognitive state. Suppose, for concreteness, that it is *disapproving of killing*. As with (H), since the expressed state is not representational, this correctness condition does not ground a representational content for (E).

In short: pure expressivism begins with the view that there is a uniform expression relation that explains the meanings of both ethical and non-ethical utterances by appeal to expression relations between sentence utterances and mental states. It

---

[14] This proposal follows (Schroeder 2008b). Notice that one *might* think that the correct assertability condition for a sentence is its truth condition. But this proposal does not make that assumption. It should be noted that some philosophers deny that Expression is crucial for developing expressivism. For example, see (Yalcin 2018). It is beyond the scope of this paper to delve into the debates about proposed alternative ways of developing expressivism without Expression.

[15] The precise correctness conditions are controversial; they might require knowing that water is wet, or justifiably believing that water is wet, etc. Here, for simplicity, we ignore this debate.

then suggests that utterance of non-ethical sentences like (S) express beliefs, while uttering an ethical sentence like (E) expresses a non-cognitive attitude.[16]

Expressivism is often motivated and developed in opposition to certain key foils. Consider two such foils, characterized within the mentalist semantic framework.

The first is *invariantist descriptivism*. Put roughly, this is the view that the semantics of ethical sentences like (E) are just like the semantics of ordinary descriptive sentences like (S). That is, (E) expresses the belief *that it is wrong to kill*, and has the representational content associated with that belief.

The second is *speaker subjectivism*. This view claims that ethical sentences like (E) express *the belief* that one has a certain non-cognitive attitude, rather than directly expressing that attitude. On this view, (E) expresses the belief that *[the speaker] disapproves of killing*. And (E) thus represents that fact about the speaker's psychology. This contrasts with the pure expressivist's account, according to which (E) does not represent anything.

There is significant debate about whether (and to what extent) the contrast between expressivism and these foils can be vindicated.[17] However, these contrasts are important for the (purported) distinctiveness of (and motivations for) expressivism.[18] So, in keeping our general argumentative strategy in this paper, we grant that these contrasts are sound.

### 1.2 Conciliatory expressivism

Pure expressivism leaves open a range of important questions about ethical thought and talk. Importantly, for our purposes, it leaves open whether we can felicitously speak of ethical "beliefs", "truths", "facts", or "properties".

One natural way of developing expressivism entails that it is *not* felicitous to speak of ethical beliefs, truths (or falsehoods), properties, or facts.[19] For reasons that will become clear, we dub this view *non-conciliatory* expressivism. Consider a sketch of how one might argue for this view.

First, as we hinted in §1.1, a natural functionalist view of belief treats belief as having the cognitive direction of fit. Given that expressivism embraces non-cognitivism about ethical thought, the natural view entails that statements like "I believe that it is wrong to kill" are infelicitous. Compare: one can't believe that "hooray for Coco". One can, of course, believe *that I disapprove of causing pain*. On simple speaker-

[16] It is controversial whether the expressivist can provide a satisfactory compositional semantics. For systematic attempts, see e.g. (Gibbard 2003) and (Schroeder 2008a). As elsewhere, here we grant the expressivist this assumption.
[17] See, e.g., (Jackson and Pettit 1998).
[18] For discussion, see (Schroeder 2014).
[19] This non-conciliatory upshot is naturally suggested by the early version of expressivism in (Ayer 1936/1971).

subjectivism, this is an ethical belief. But the expressivist denies that this is an *ethical* belief: it is merely a belief about one's own psychology.

Second, given mentalism, it is natural to identify the truth conditions for a sentence with the content of the belief expressed by that sentence. For example, for sentence (S) to be true just is for water to be wet. By contrast, (H) does not express a belief. This explains why (H) cannot be true or false. The non-conciliatory expressivist concludes that ethical sentences like (E) cannot be true or false, for the same reason.

Third, it is natural to think that *it is a fact that p* only if *it is true that p*. Because the non-conciliatory expressivist denies that there are ethical truths, they also deny that there are ethical facts.

Finally, consider the predicate expression "…is wet" in (S). Because of the representational direction of fit of the thought *water is wet*, the embedded predicate expression is naturally understood as representing the property of being wet. But now consider the predicate expression "…is wrong" in (E). According to pure expressivism, (E) does not function to represent. So, according to the non-conciliatory expressivist, talk of a "property" of wrongness is nonsense.

Non-conciliatory expressivism is a natural and well-motivated way to develop pure expressivism. But many philosophers take it to be implausible or even hopeless as a metaethical view. To see why, consider the semantic data provided by sentences like "it is true that cruelty is wrong"; "it is a fact that cruelty is not wrong"; "I believe that cruelty is sometimes wrong"; "badness is a monadic property, but wrongness is a relation"; etc. Each of these sentences are apparently each *perfectly semantically felicitous* to utter. This contrasts with analogous sentences based on (H): it is *not* felicitous to utter "it is true that hooray for Coco", etc. So non-conciliatory expressivism seems to make strikingly incorrect predictions about the felicity conditions for utterances involving ethical terms, including many ones that ordinary speakers utter in everyday circumstances. Because expressivism is a descriptive theory about our actual language, such incorrect predictions appear damning.[20]

This motivates a project that Sebastian Köhler has recently usefully dubbed *conciliatory expressivism*.[21] This project aims to make expressivism compatible with the felicitous assertability of the sorts of sentences mentioned in the previous paragraph. In principle, the expressivist might seek to achieve "conciliation" with the felicity data in a variety of ways. In practice, however, (pure) expressivists have typically sought to achieve conciliation by appealing to two sorts of resources: 1) "minimalist" semantics for certain terms and 2) expressivist semantics for certain terms. We introduce each in turn. For ease of exposition, in the core of this paper, we assume that these two resources exhaust those that the expressivist can use to

[20] For this reason, some see non-conciliatory expressivism as a non-starter. See (Cuneo 2013, 223), (Dreier 2010, 170), and (Price 2015, §1.2). For resources to defend the non-conciliatory expressivist, see (McPherson 2022).
[21] (Köhler 2021).

achieve conciliation. We return in the conclusion to the possibility that there might be other resources that could be used to achieve conciliation.

To illustrate minimalism, consider the case of 'true'. On a simple version of minimalism about 'true', the meaning of the word 'true' is exhausted by the fact that for every well-formed indicative sentence 'S', 'S' is true just in case S. Because both (S) and (E) are indicative sentences, they are each truth-apt, according to minimalism, while (H) is not an indicative sentence, and so is not truth-apt.[22]

Crucially, minimalism makes truth-aptness depend on a syntactic property: namely, the property of being *indicative*. This property is wholly independent of the mentalistic semantic picture offered by the expressivist. So, if minimalism is correct, the non-conciliatory expressivist's rationale for denying the felicitous application of 'true' to ethical sentences is undercut.

The minimalist strategy can be extended to terms like 'fact', 'belief, and 'property'. For example, one might claim that the meaning of 'fact' is exhausted by the fact that for every well-formed indicative sentence 'S', if 'S' is true, then it is a fact that S. One might claim that the meaning of 'belief' is exhausted by the fact that for every well-formed indicative sentence 'S', belief that S is whatever mental state is semantically expressed by 'S'. And one might claim that the meaning of 'property' is exhausted by the fact that every well-formed sentence of the form 'x has the property of being F' is equivalent in meaning to 'x is F.'[23] If these sorts of minimalist hypotheses are correct about our actual language, then the expressivist can accommodate much of the felicity data mentioned above.

Now consider the second kind of conciliatory resource we mentioned earlier: expressivist accounts of additional parts of thought and talk. For example, consider an expressivist hypothesis about the word 'objectively'.

To illustrate the main idea, imagine the possibility of waking up tomorrow enjoying the taste of something you currently find disgusting (say: durian), and also approving of torturing children. Intuitively, you might be sanguine about the first possibility: "if I were to enjoy eating durian, great!" But you might react to the second possibility with horror: "torturing children would still be awful, even if I came to approve of it!" Our expressivist hypothesis suggests that ethical 'objectivity' talk expresses the second sort of attitude:

| | |
|---|---|
| **'Objectivity' expressivism** | 'X is objectively wrong' expresses disapproval of X that is insensitive to counterfactual attitudes, e.g. disapproval even for the circumstance where I and everyone else approve of X.[24] |

[22] Minimalism should be distinguished from the broader idea of offering a "deflationary" account of metaphysical vocabulary. For an example of the latter sort of project that doesn't rely on minimalism, see (Kraut 1993).
[23] For discussion of this general strategy see (Dreier 2004), (McPherson 2022) and (McPherson 2019a).
[24] Compare variants of this idea in (Blackburn 1998) and (Gibbard 2003).

'Objectivity' expressivism isn't a minimalist semantic hypothesis. Rather, it is a substantive expressivist analysis of the meaning of the relevant sentence-type.

A successful conciliatory expressivist project would use such minimalist and/or expressivist analyses to secure the semantic felicity of asserting a range of ordinary sentences involving ethical vocabulary. Crucially, felicity is different from accuracy. To illustrate: it is felicitous, but inaccurate, to say that wearing pants is illegal in Ohio. This means that a conciliatory expressivist could either accept a sentence like "torturing children is objectively wrong", deny it, or remain uncommitted about it. Indeed, in theory, someone could endorse conciliatory expressivism while being wholly ethically uncommitted.[25]

What Köhler dubs "quasi-realism proper" combines conciliatory expressivism with further claims that are often taken to be associated with "realist" metaethical views.[26] For example, quasi-realists characteristically accept claims like "torturing children is objectively wrong".[27] In so doing, they accept forms of ethical objectivity and "attitude-independence" often considered to be a mark of realism.[28] Most conciliatory expressivists are quasi-realists. But Köhler's formulation also allows us to characterize a sort of "quasi-anti-realist": namely, a conciliatory expressivist who rejects the above sort of ethical "objectivity" claims. Note that "quasi-realism" and "quasi-anti-realism" are simple extremes: a conciliatory expressivist could instead accept more complicated claims about the kinds (and extent) of objectivity involved in ethics.

In what follows, we chiefly address the ethically committed "quasi-realist" conciliatory expressivist. But in considering the explanatory challenges presented in the introduction, we need to clearly understand the contrasting significance of conciliatory hypotheses about language, on the one hand, and the quasi-realist commitments, on the other.

## 2. Explanatory Challenges, and How to Undercut or Answer Them

We now more carefully introduce the two explanatory challenges that are our focus: the supervenience and reliability challenges. In this section, we explain each in turn. We then distinguish the two fundamental ways of responding to such challenges: *undercutting* and *answering*.

### 2.1 The supervenience challenge

[25] This is a point that Sebastian Köhler has helpfully underscored in recent work. See (Köhler 2021). However, see (Faraci 2017) for a challenge to whether the expressivist can be uncommitted in this way.
[26] In the literature, it is often unclear whether 'quasi-realism' is being used to pick out what Köhler calls "conciliatory expressivism", what he calls "quasi-realism proper", or both. We think the distinction that Köhler is drawing attention to is crucial. For the purposes of this paper, we follow him in marking it using his preferred terminology for doing so.
[27] For example, see (Blackburn 1998) and (Gibbard 2003).
[28] It is controversial whether and how quasi-realist expressivism can be meaningfully distinguished from the standard "realist" views it mimics. See (Dreier 2004) for discussion. We again grant the conciliatory expressivist that their view is distinct.

The supervenience challenge arises from a purported datum: ethical features cannot vary independently of certain other features. For concreteness, we will use the following common formulation of this datum:

**Supervenience** If two metaphysically possible entities (e.g. actions, states of affairs, persons…) are alike in all natural respects, they are alike in all ethical respects.

This is a claim about the relation between two classes of *properties*: the ethical properties and the natural properties.[29] It states that the pattern of instantiation of the first class of properties cannot vary across possible circumstances without some variation in the pattern of instantiation of the second class. Importantly, Supervenience says that this is true across *all metaphysically possible worlds*.[30] It is not, for example, an idiosyncratic fact about our world, or a pattern we only find in "normal" or "nearby" worlds.

In recent work, Gideon Rosen claims that this sort of supervenience thesis might be the "least controversial thesis in metaethics".[31] To get a feel for the plausibility of Supervenience, consider an act from your own life that you are certain was wrong.[32] Now hold fixed *absolutely everything* non-ethical about the world: saliently, hold fixed the wrongful actor's history and mental states, their intentions, the consequences of their action, etc. Now try to imagine a world in which all of that is held fixed, and yet their act was right rather than wrong. This seems impossible. But there is nothing special about that specific case; this sort of reasoning would seemingly be compelling for any case you thought carefully about.

The supervenience *challenge* is to explain Supervenience. Different metaethical views appear to fare better or worse with respect to this challenge. For example, one kind of naturalistic realism claims that each ethical property is identical to some natural property. This ensures that every ethical difference will be a natural difference, thus securing Supervenience. Other realist views seem to have more difficulty in providing the relevant explanation. For example, on common forms of non-naturalism, the ethical properties are radically "different in kind" from the natural properties, such that it is controversial whether advocates of this view can explain Supervenience.[33]

### 2.2 The reliability challenge

[29] There are reasons to worry about understanding the second class of relata as "natural". See e.g., (Sturgeon 2009). For a more complex formulation intended to address these worries, see (McPherson 2012, 213-214). We ignore these complications, since they do not matter in the context of this paper.

[30] For the purposes of this paper, we are neutral on how to understand the metaphysics of "possible worlds" as such. We also don't delve into exactly what makes a possible world "metaphysically" possible in particular.

[31] (Rosen 2020, 206). Rosen goes on to argue against Supervenience.

[32] For a more detailed and careful motivation, see (McPherson 2019b, §2).

[33] For discussion, see (McPherson 2012) and (McPherson and Plunkett 2025b).

Now consider the "reliability challenge".[34] This challenge begins with a modest anti-skeptical hypothesis:

**Reliability** An identifiable (and important) subset of our actual ethical beliefs reliably tracks the ethical truths.

One might motivate this idea by considering that certain ethical matters seem "obvious", and focusing on those. For example, one might think that we are at least more likely to have true ethical beliefs about the badness of torturing innocents for fun than we would have been if we had acquired these beliefs at random. This suggests that these "obvious" ethical beliefs track the ethical truths with some degree of reliability.

Reliability claims *at least* entail certain modal correlations. (On some views they might just *amount to* certain modal correlation claims.) For example, for your beliefs about what's in your fridge to reliably track what's in your fridge, the content of these beliefs must "match" the fridge facts across some significant proportion of the relevant "nearby" possible worlds.

It is worth noting here that the correlation at issue in Reliability matters for further claims many want to make about the epistemology of ethics. For example, a relevant sort of reliability is arguably required for ethical knowledge.

The reliability *challenge* is to explain Reliability. As with Supervenience, some forms of realism seem to smoothly enable such an explanation (if they are true). The most straightforward such explanations appeal to either causal or constitutive relations. Consider an example of each.

Suppose that the ethical facts are causally efficacious. If they are, then the ethical facts might affect our perceptual capacities in a way that enables us to track these facts. By analogy, a causal explanation would be a promising explanation of the reliability of your beliefs about what's in your fridge. Or suppose instead that the ethical facts are partly *constituted* by our ethical beliefs. (On the simplest such view, the fact that it is wrong for S to do A just is the fact that S believes that it is wrong to do A.) This sort of constitutive claim, if true, would also explain the reliable correlation between many ethical beliefs and the relevant ethical facts.

As with the supervenience challenge, however, some forms of realism appear to have more difficulty explaining Reliability. For example, advocates of standard non-naturalist forms of metaethical realism deny that ethical properties are causally efficacious, or that they are constituted by ethical beliefs. This cuts off their ability to explain Reliability in the straightforward ways that other metaethical views can.

[34] This label for this sort of challenge comes from (Schechter 2010). See (Schechter 2017) for an overview of this kind of challenge in metaethics, which we draw on here. See also (Enoch 2011) and (Plunkett 2020).

It is controversial how, or whether, non-naturalistic realist views can explain Reliability in other ways.[35]

### 2.3 Three clarifications about these challenges

Note three clarifications about these challenges and their significance, which are important for our arguments to come.

First, for each challenge, at least some (if not all) of the items that are correlated are "object-level" rather than "representational-level" entities. For the supervenience challenge, the issue is explaining the supervenience of the ethical facts themselves on the natural facts. It is thus a purely object-level challenge. The reliability challenge concerns correlations between "representational-level" facts (facts about our ethical thought and talk, such as what the content of such thought and talk is) and object-level facts (facts about which ethical facts actually obtain).

Second, a satisfactory answer to one of these challenges need not entail a satisfactory answer to the other. For example, the thesis that the ethical facts are naturalistic might underwrite a good answer to the Supervenience challenge, but leave the reliability challenge unanswered (because we do not reliably track all naturalistic facts).

Third, there are important controversies concerning how to formulate the two challenges. We have generally elided these in our discussion. This is partly for brevity. But it is mostly because we hope to show that our core line of argument, about the explanatory limitations of conciliatory expressivism with respect to the supervenience and reliability challenges, is largely independent of further details about how to formulate these challenges – provided that we hold fixed the crucial idea that both challenges involve (at least some) "object-level" ethical relata.

### 2.4 Undercutting versus answering

One controversy about the supervenience and reliability challenges concerns how dialectically significant they are. Our own view is that it is very important for a metaethical view to be able to address these challenges.[36] However, we do not argue that here, because we take it to be common ground with our foil. Recall from the introduction that our foil takes it to be a significant theoretical virtue of expressivism that it has distinctive resources that enable it to address these challenges.

How might the conciliatory expressivist's resources address these challenges? If one takes these challenges to be dialectically significant, one can successfully address them in either of two very different ways. One can *undercut* the alleged applicability

[35] See (Schechter 2017) for an overview.
[36] For some of our relevant discussion of these challenges, see (McPherson 2012), (Plunkett 2020), and (McPherson and Plunkett 2025b).

of the challenges. Or one can accept the challenges as legitimate and *answer* them. We introduce these sorts of replies in turn.

The basic way to undercut a challenge is to claim that the challenge is mistaken or confused. Because we understand each of our target challenges as involving the demand to explain a correlation, these challenges can be undercut by denying the existence of (at least) one of the relata of the alleged correlation.

To see how this might be done, consider non-conciliatory expressivism, introduced in §1.2. On a thoroughly non-conciliatory version of this view, talk of ethical "properties", "belief", and "truth" talk is all infelicitous. This view might seem to suggest that there is no non-defective way of stating the relevant correlations at the heart of the challenges: e.g., those involving "ethical properties" (as suggested by Supervenience), or "ethical truths" and "ethical beliefs" (as suggested by Reliability).

To see the force of the undercutting strategy, imagine a philosopher who notices some sports fans cheering. She then sets out to explain the correlations between the "hooray beliefs" (that she supposes to be expressed by the cheers) and the "hooray facts" (that the "hooray beliefs" are supposedly about). On the standard picture of cheering, this explanatory project is profoundly misguided: there simply are no "hooray beliefs" or "hooray facts", of the kinds imagined by this philosopher.

It's important to emphasize that we introduce this undercutting strategy as an illustration of how these challenges *might* be undercut, using resources from non-conciliatory expressivism in particular, rather than as a claim about what all advocates of non-conciliatory expressivism can in fact accomplish. Indeed, given what we go on to argue in the rest of this paper, many forms of non-conciliatory expressivism might be unable to undercut one both of these challenges in this way. In short, this is because some forms of non-conciliatory expressivism will, in fact, give us the resources to state the relevant correlations at the heart of one (or both) of these challenges.[37]

Turn next to *answering* these challenges. Unlike undercutting, this strategy grants that these are genuine, and not merely apparent, explanatory challenges. To answer such a challenge is to offer a satisfactory answer to the central explanatory question. In our cases, answering thus involves granting that the correlations stated in Supervenience and Reliability obtain, and then explaining why these correlations hold (and, perhaps, illuminating something more about their status).

[37] When thinking about this result, it is worth keeping in mind three things. First, different versions of expressivism can be more or less conciliatory, and be more or less so with regard to different fragments of thought and talk. (For example: some might conciliate to a high degree on "belief" talk, but much less so on "truth" talk, while others vice versa.) Second, the supervenience and reliability challenges involve different correlations. (This means that it might be possible, given a certain form of non-conciliatory expressivism, to state the relevant correlation at the heart of one of the challenges, but not the other.) Third, as we argue more below, the supervenience and reliability challenges can each be stated with fewer resources than is often assumed. (For example, as we argue below in §3, we don't actually need "property" talk to state the challenges).

In §§2.1-2.2 we briefly gestured at theories that are promising for answering each of the relevant challenges. Let's now work through an example slightly more carefully. Recall the speaker subjectivism briefly introduced in §1.1 above. On a simple way of developing this view, wrongness is identical to a relational property: the property of being disapproved of by the speaker. And so, on this view, for a speaker S's belief *that x is wrong* to be true is just for S to disapprove of x.

This kind of view (if true) would smoothly explain both Supervenience and Reliability. Supervenience is explained because, on this view, the ethical features (such as wrongness) are among the natural features. So the correlation between the ethical and the natural is explained trivially. On simple subjectivism, the relevant wrongness facts just are facts about our (dis)approvals. So, for Reliability to be explained, we would need to explain a correlation between our beliefs about wrongness and our (dis)approvals. This correlation can be explained by psychological mechanisms that tend to reliably align these parts of our psychology, whose existence can in turn be potentially explained by the evolutionary benefits of this belief-attitude correlation among our ancestors.

We take "undercutting" and "answering" to exhaust the broad ways that one can address an explanatory challenge to one's view. So, we can clarify the dialectic: for the conciliatory expressivist to have distinctive resources that enable them to address these challenges, these resources must enable them to either undercut or to answer the challenges. We now argue that the conciliatory expressivist has no such resources, considering undercutting and answering in turn.

## 3. Conciliatory Expressivism Cannot Undercut the Challenges

In this section, we consider whether the conciliatory expressivist can offer an undercutting response. Our focus is on views that are (at least) *minimally conciliatory*, in the sense that they do not require that we reject as infelicitous a wide range of seemingly obviously felicitous claims. As we noted earlier, non-conciliatory expressivism seemed to undercut these challenges in a straightforward way: if true, it made it infelicitous to state these challenges. We first argue that, by contrast, expressivist views that achieve (at least) minimal conciliation *do* enable us to felicitously state the challenges. This leaves open the possibility that the conciliatory expressivist has some *other* way of undercutting the challenges. We consider three strategies for doing so, but argue that none of these strategies are promising.

### 3.1 Felicitously stating the challenges

The conciliatory expressivist cannot undercut these challenges in the way that the non-conciliatory expressivist can. To see why, focus on predication. Above, we offered a simple toy expressivist gloss on "ethical predication", according to which simple sentences of the form 'x is wrong' express *disapproval of x*. But this is far from an adequate expressivist semantics for the predicate 'wrong'. If something is a

genuine *predicate*, it needs to *behave the way a predicate behaves*, across a wide variety of contexts.[38] Expressivists have famously wrestled with one part of this challenge: the "embedding" of predicates, such as in the antecedents of conditionals. Consider: "if Alice is good, then Sam will vote for Alice." This sentence does *not* express approval of Alice. Attempting to give a satisfactory expressivist semantics for this sort of sentence is at the heart of the so-called "Frege-Geach" problem for expressivists.[39]

Here's a different datum about how predicates behave: they are sufficient to enable the smooth statement of *correlations*. Quite generally, we can felicitously use any pair of predicates 'F' and 'G' to state that "Fness" and "Gness" walk around together. And, again quite generally, these correlation statements embed smoothly in modal contexts. This is also true of *ethical* predicates in particular. For example, it is felicitous to state alleged ethical modal correlations like "most possible actions that are motivated by bloodlust are wrong". We thus think that in order to be minimally conciliatory, expressivism needs to enable us to state modal correlations.

Next, notice that a minimally conciliatory version of expressivism will also enable us to felicitously ask (and propose answers to) *explanatory questions* about ethical correlation claims. For example, one can felicitously ask the question of why paternalism towards competent adults is ordinarily wrong, and then answer it by saying it is *because* it fails to respect their agency.

Putting these observations together, a minimally conciliatory expressivism needs to enable us to do two things. First, it must enable us to felicitously *state* modal correlations (such as Supervenience and Reliability). Second, it must enable us to felicitously ask (and propose answers to) explanatory questions about those correlations. Together, these requirements strongly suggest that the conciliatory expressivist cannot undercut the explanatory challenges in the way that the simplest non-conciliatory expressivist does.[40]

We cannot show that there is no possible conciliatory expressivist account that undercuts. But we can offer general reasons for pessimism. From §1, we know that if an adequately conciliatory account can be offered, it will need to appeal to either minimalist or expressivist interpretations of certain crucial pieces of vocabulary. In what follows, we will illustrate the limitations of these resources by considering an undercutting attempt that appeals to the minimalist character of relevant vocabulary, and one which appeals to the expressivist character of relevant vocabulary. We begin, however, by considering an ingenious argument that we have reason to believe the explanatory challenges are undercut, even if we cannot spell out the details of the undercutting attempt.

---

[38] We are thus suspicious of Schechter's idea that we could introduce a semantically undefective predicate that *simply* functions to state a command. (See (Schechter Manuscript).) For a general statement of the problems faced by this idea, see (Dreier 1996).

[39] For an overview, see (Woods 2017).

[40] It is not obvious that an expressivist can explain the felicity of the correlation and explanatory claims we have emphasized. Here, we follow our usual strategy throughout this paper: we grant, for the sake of argument, that the expressivist can offer such an explanation.

### 3.2 Dreier's case for undercutting

Jamie Dreier has offered an interesting argument for what we are calling the "undercutting" response to the reliability challenge.[41] Simplified, Dreier's core strategy is this. Begin with a language that is as A. J. Ayer imagines our language to be. (Ayer's view is very similar to what we dubbed "non-conciliatory expressivism" in §1.2). Dreier notes that Ayer's hypothesis appears to vindicate undercutting, since the relevant language simply lacks the expressive power to state the correlations that underlie the challenge. Dreier then imagines enriching the Ayer-style language in a stepwise fashion, introducing (expressivist-interpreted) ethical predicates, then minimal truth predicates, and then "objectivity" talk. At each step, Dreier argues that if the reliability challenge did not arise at the previous step, nothing in *this* step enables it to arise.[42]

Dreier grants that there is something unsatisfying in his argument. For he argues that the standard conciliatory expressivist accepts a collection of claims that seem jointly sufficient for posing the reliability challenge.[43] He suggests that his stepwise argument shows that these apparently jointly sufficient conditions cannot in fact be sufficient, without explaining why they fail to be so.

Contrary to Dreier, we think that it is possible to extract from his discussion a proposed explanation for why conciliatory expressivism can still undercut the reliability challenge. This proposed explanation begins by noting that the reliability challenges is not about language at all: it is a challenge about the relationship between mental states and "object-level" states. The proposed explanation then suggests that, at each step in Dreier's enrichment, we don't add new *mental states* beyond those already expressible by an Ayer-style language. Rather, we simply add new resources for *conventionally linguistically expressing* mental states that speakers of an Ayer-style language could already possibly have. But if the Ayer-style account enabled us to undercut the challenges, simply adding expressive power to its *language* should not resurrect the reliability challenge.[44]

We also think that this clarification enables us to see where Dreier's argument goes awry. Start with a non-cognitive mental state type: say, *disapproval*. Ayer's view

---

[41] (Dreier 2012).

[42] Strikingly, Dreier forcefully prosecutes the supervenience challenge to the conciliatory expressivist in (Dreier 2015), where (put roughly) he argues that they cannot undercut the challenge, and lack a good answer to it. As we flagged earlier, our argument in this paper owes much to Dreier's discussion. We think there is a significant tension between that discussion of the supervenience challenge, and his discussion of the reliability challenge in (Dreier 2012). In short: if the conciliatory expressivists can undercut the reliability challenge in the way Dreier suggests in (Dreier 2012), this strategy will generalize to undercut many other explanatory challenges, including the supervenience challenge.

[43] (Dreier 2012, 286).

[44] Another possibility is that the relevant Ayer-style account involves a psychology that enables us to think correlations, but the semantics does not enable us to linguistically express them. We set this interesting possibility aside. It should be noted that, if this possibility were true, it would frustrate, rather than help, someone who wanted to appeal to continuity with an Ayer-style account to undercut our explanatory challenges.

starts with the idea that a person can disapprove of certain actions. Now we imagine introducing the novel predicate 'is wrong'. Simple sentences of the form 'x is wrong' express *disapproval of x*. You might think that there are no new mental states here, just a new linguistic vehicle. But as we argued in §3.1, anything that behaves semantically like genuine predication will enable us to felicitously state modal correlations. So, if we grant that Ayer's account of ethical thought didn't enable us to state such correlations, then successfully introducing anything that behaves semantically like predication *presupposes* that there are mental states (ethical "correlation thoughts") that (we are granting) Ayer's account rules out.

Dreier notes that someone might be suspicious of the expressivist's ability to successfully introduce predicates. But he claims that this is a separate issue, unrelated to the reliability challenge.[45] However, our point is not that the expressivist cannot introduce predicates. It's true that one possibility is that there is no non-cognitive mental state that can be expressed with a genuine predicate.[46] If that's true, then so much the worse for the conciliatory expressivist's project. Our point is different. It is that part of successfully introducing a predicate expression is, inter alia, enabling felicitous statement of correlations. So if the expressivist *can* introduce ethical predicates, then they enable the felicitous statement of correlations, and hence the felicitous posing of the reliability challenge.[47]

### 3.3 Undercutting via minimalism?

How might conciliatory expressivists who aspire to undercutting respond? A natural strategy appeals to the following idea: because of a commitment to minimalist semantics, the kinds of "facts", "truths", "beliefs", "properties" etc. that conciliatory expressivism vindicates talking about aren't of the *same kind* as those required to generate the relevant explanatory challenges.[48]

They might develop this thought in the following way. Conciliatory expressivism proceeds by providing "minimalist" characterizations of metaphysical-sounding words like 'true', fact', etc. But surely the explanatory challenges instead require that the relata be "metaphysically heavyweight" Truths, Facts, etc. So, while it appears that the challenges can be stated, this is an illusion caused by the "surface level" features of our thought and talk. Understanding the consequences of the minimalist interpretation of the relevant vocabulary suffices to undercut the challenges.

This reply fails.[49] The crucial point is that, although the supervenience and reliability challenges are each smoothly (and perhaps most naturally) stated in terms of "truths", "facts", etc., the challenges themselves don't need to be put in this way.

---

[45] (Dreier 2012, 280).
[46] For connected discussion, see (Finlay Manuscript).
[47] It should be underscored that this claim about predication is about what's sufficient for stating the challenges, and not about what is necessary for doing so. It's an interesting further question about exactly which linguistic resources are necessary for stating the challenges, but we don't take a stand on that here. See (Schechter Manuscript) for connected discussion.
[48] Compare (Gibbard 2011).
[49] For connected discussion, see (Golub 2017a) and (Schechter Manuscript).

Instead, as we have emphasized, each requires only that we be able to state certain *correlations*. For example, consider a supervenience thesis: no two possible actions can differ with respect to the property of wrongness, without also differing in *some natural property*. We have property talk on both sides of this thesis. But all that we need from such talk are the usual "minimalist" virtues of expressive convenience. For example, suppose that we had a complete list of the natural predicates and relations, [F, G,….]. Then we could rewrite our thesis as: no two possible actions can differ in wrongness, without also differing in at least one of Fness, Gness…. So the 'property' talk here is in principle dispensable.[50]

The broader lesson of this discussion is that talk of "truths", "facts", etc. is not essential to stating the explanatory challenges we are focused on. Because of this, it is implausible that minimalist theories of these words can be used to undercut the challenges. (In §4, we will consider the separate question of whether minimalist resources can be used to *answer* the challenges.)

### 3.4 Golub's functional role argument

The discussion just concluded suggests that appeal to the conciliatory expressivist's *minimalist* semantic resources is not helpful for undercutting the explanatory challenges. Now consider an attempt to appeal to distinctively *expressivist* resources.

Camil Golub suggests that the conciliatory expressivist can undercut the reliability challenge by appealing to part of the non-cognitivist's thesis about the functional role of ethical thoughts.[51] In the terms we have introduced, such thoughts do not have the "belief-like" world-to-mind direction of fit.[52]

Golub proposes that "we should not demand a certain kind of explanation of our reliable access to normative facts, because this would saddle normative thought with a function that it does not have."[53] He suggests that the conciliatory expressivist can offer a sort of explanation of reliability: one that (he claims) any committed ethical inquirer could offer. He suggests that call for further explanation of reliability is apt for metaethical "representationalists", for whom ethical thought has a "tracking function", but not for the conciliatory expressivist. We can sum up this view as follows:

| | |
|---|---|
| **Functional role-reliability link** | Aptness of the reliability challenge for a fragment of thought is connected to a "representational functional role". |

In our view, the explanatory relationship between mental state function and epistemological norms is left unsatisfyingly elusive in Golub's discussion. And

[50] Schechter makes a parallel point about the reliability challenge in (Schechter Manuscript).
[51] (Golub 2017b).
[52] (Golub 2017b, 804) talks about a "naturally representational" function, which is arguably a bit less committal than our "direction of fit" talk. This doesn't make a difference to our argument below.
[53] (Golub 2017b, 800).

extensionally, it is not obvious to us why the relevant functional claims undercut some epistemological challenges ("a certain kind"), but apparently leave others untouched.[54] We can see why Golub's proposal is unsatisfying by considering two alleged rational links:

| | |
|---|---|
| **Belief-reliability link** | It is irrational to believe both (a) that it is wrong to do A and (b) that belief (a) is highly unreliable. |
| **Reliability-explanation link** | It is irrational to believe both (c) that one's own ethical beliefs reliably track the ethical facts and (d) that there is no satisfactory explanation for belief (c). |

We treat "satisfactory explanation" as a placeholder here, for the explanatory condition required to meet the reliability challenge. The nature of this condition is a matter of live controversy within that literature. On one natural way of thinking about the reliability challenge, it rests on the Reliability-explanation link (or something closely related).

Here is our puzzle for Golub. Golub seems to want to use the non-representational function only to reject Reliability-explanation link, and not Belief-reliability link. In one sense, this is understandable: rejecting Belief-reliability link would be strikingly revisionary. But we don't think that non-cognitivist functionalism provides the resources to reject only Reliability-explanation link and not Belief-reliability link. Suppose, as Golub appears to suggest, that the epistemic norms on ethical beliefs are explained by the alleged non-representational function of ethical beliefs. The most natural upshot of this would be to reject Belief-reliability link. The core thought would be that non-representationality severs the rational connection between reliable tracking and belief.

In order to reply to this, on behalf of Golub's line of thought, one might seek to distinguish "internal" reliability judgments from "representational" ones. For example, we could spell out an "internal" reliability worry for an agent as one about whether she would continue to have the target beliefs if she were ideally coherent.[55] By contrast, "representational" reliability worries would be tied to "representational accuracy", which in some sense "reaches outside" of her psychology. One might claim that only the cognitivist faces a "representational" version of the Reliability-explanation link.

There are at least two worries here. The first is that this proposal makes it hard for the conciliatory expressivist to accommodate felicitous "fundamental moral error"

[54] Some views about the foundations of epistemology take epistemic norms to be (or be grounded in) the constitutive norms for cognitive states. On these views, epistemic norms would be simply inapplicable to non-cognitive states. This would challenge the *conciliatory* credentials of conciliatory expressivism.

[55] Compare the account of one's fallibility given in (Blackburn 1998, 318).

thoughts.[56] For example, it seems like one can coherently worry that one's ethical opinions might be badly off-track *even if they were made maximally coherent with each other*. The second worry is that this contrast is most smoothly vindicated by a form of *quasi-anti-realism* on which the correctness conditions on one's ethical thoughts are given by idealizing one's psychology.[57] For the person who embraces quasi-anti-realism, "fundamental moral error" worries are coherent, but simply substantively misguided. Absent this "anti-realist" sort of commitment, it is unclear why the expressivist must be worried about "internal" unreliability. Why couldn't an agent be rationally indifferent to it, given the non-representational function of her ethical judgments?

The deeper point is that, on this picture, the quasi-anti-realist's metaethically optional substantive normative commitments are doing all the crucial work, and the conciliatory expressivism is doing none. To see this, note that a cognitivist with the same anti-realist commitments can address the challenge in exactly the same way.

Our discussion casts doubt on Golub's proposal. But it does more than that: it also casts doubt on the prospects for appealing to non-cognitive "function" to undercut the explanatory challenges. This is because it illustrates two general facts. First, the most natural way of taking non-cognitivism to have epistemic implications will devastate the expressivist's conciliatory ambitions. Second, the expressivist's most promising way of replying to the challenges owes nothing to the expressivist's own distinctive commitments. We will see this pattern repeat going forward.

### 3.5 Summing up

We have not exhaustively discussed every attempt to undercut the explanatory challenges. However, we take our discussion to show that it is unclear how conciliatory expressivism could undercut them. This is for three general reasons. First, fewer resources are needed to felicitously pose the challenges than some conciliatory expressivists have appreciated; we suggested that (recognizable) ethical predication and modal relations suffice. Second, appealing to *minimalist* resources to undercut the challenges is misguided. And, third, appeals to distinctively *expressivist* resources face a difficult challenge: to vindicate enough of our felicity judgments to retain a plausibly conciliatory hypothesis, while somehow undercutting just the challenges that we are targeting, in a way that the realist cannot also adopt.

## 4. Conciliatory Expressivism Cannot Answer the Challenges

We've argued that conciliatory expressivists can't *undercut* the supervenience and reliability challenges. We now argue that the conciliatory expressivist cannot *answer* these challenges in a way that appeals to distinctive resources. That is, where the conciliatory expressivist may be able to answer these challenges, their ability to do

[56] Cf. (Egan 2007).
[57] Compare (Street 2011).

so depends wholly upon auxiliary theses that many realists can (and often do) endorse.[58] As with the discussion of undercutting, we argue that the problem arises from a mismatch between the nature of the challenges, and the kinds of resources that conciliatory expressivists can bring to bear on them.

For simplicity, we focus in this section on Supervenience. We begin by distinguishing our explanandum from related theses about rational commitment. We then discuss a central sort of strategy for explaining Supervenience by appeal to normative claims. We argue that neither minimalist nor expressivist resources help the expressivist offer a better version of this style of explanation.

### 4.1. Contrasting explananda

It might seem that the expressivist can elegantly explain Supervenience by appealing to distinctive expressivist resources. For example, Allan Gibbard develops an account of ethical thought that appeals to "hyperplans", where these are, roughly, plans that include a response to every conceivable natural contingency. Gibbard argues that, on his account, anyone who treats apparently naturally identical circumstances as ethically different will have plans that cannot be "filled out" to form a consistent hyperplan.[59] Explaining this kind of commitment to Supervenience (or some related feature of our thought and talk) is all that some expressivists aspire to explain. But, in the context of our discussion, the key question is whether explaining this commitment (assuming it can be done) can, in turn, be used to address the supervenience challenge we've been discussing. And, here, one might be tempted to think the following: the kind of inconsistency Gibbard identifies *can* be used to answer the supervenience challenge.[60]

Many philosophers think that the supervenience of the ethical is a kind of a priori truth, guaranteed to be true by the nature of our ethical concepts.[61] Gibbard purports to vindicate something like this as well: we are committed to Supervenience simply by grasping the nature of ethical concepts as essentially "planning concepts".[62]

Can this Gibbardian thesis (about our commitment to Supervenience) be used to answer the supervenience challenge that we've been discussing? It cannot. The basic problem is that explaining why we are committed to Supervenience is different from explaining Supervenience itself.[63] That's true whether or not that commitment is in some sense a priori. To see this, recall Supervenience:

---

[58] Compare (Schechter Manuscript).
[59] (Gibbard 2003, 89-94).
[60] For important challenges to Gibbard's reasoning concerning our commitment to Supervenience, see (Chrisman 2005) and (Sturgeon 2009, 82-88).
[61] E.g., (Venesmaa 2021).
[62] (Gibbard 2003, 92).
[63] This point is made powerfully in (Dreier 2015, 289-290). See also (Schechter Manuscript, 16) for parallel points about the reliability challenge.

**Supervenience** If two metaphysically possible entities (e.g. actions, states of affairs, persons…) are alike in all natural respects, they are alike in all ethical respects.

This is an object-level metaphysical thesis about the relationship between two classes of properties: the natural and the ethical. It's *not* about our thought, talk, and "commitments". For example, it's not a descriptive claim about how we think and talk about ethical matters, nor a normative one about how we should.

Contrast this type of metaphysical thesis with:

**S-Commitment** It is rational for people to be committed to Supervenience.

This is a normative claim about our psychologies.[64]

In general, explaining *why P* is a completely different task from explaining *why we are rationally committed to P*.[65] We can illustrate this contrast by considering Cartesian "cogito" reasoning. This reasoning purports to explain why each of us is *rationally committed* to our own existence. But it does *not* purport to explain the fact that *you now exist*. (The latter fact is presumably to be explained by a series of contingent historical facts.) More generally, as "error theorists" in metaethics and other domains remind us, it could turn out that our thought and talk in a given domain commits us to there being certain kinds of facts that simply do not obtain.[66]

So even if Gibbard has provided an excellent explanation of S-Commitment, he would not thereby have explained Supervenience. This explanatory silence is not a result of some idiosyncrasy of Gibbard's view. Rather, it reflects a deep fact about the core expressivist account: namely, this is an account that appeals to distinctive claims about ethical thought and talk. It is thus no surprise that it is potentially well placed to explain certain data about thought and talk (such as the sort of commitment encapsulated in S-Commitment).[67] But the core expressivist resources are, by themselves, the wrong sorts of resources to explain object-level theses like Supervenience. (Or theses like Reliability, which involve at least some facts at the object-level.)

To be clear, one *could* appeal in part to facts about our commitments to explain Supervenience, if one accepted a different metaethical account. For example, in developing the sort of simple speaker-subjectivism we introduced earlier (in §1.1),

---

[64] Compare (Klagge 1988). One might also take the (alleged) fact that we are committed to (e.g.) Supervenience (as opposed to the rationality of that commitment) to call out for explanation. That (purported) explanandum is not our focus here.

[65] Compare (Dreier 2015).

[66] For an overview of metaethical error theory, see (Olson 2017).

[67] Some expressivists argue that the fact that they have better explanations of theses like S-Commitment is itself a reason to favor their views. See e.g. (Venesmaa 2021). We do not seek to evaluate these arguments here. Given the overall dialectic in our paper, it should be noted that there is also the question of how and why realists might themselves be able to vindicate S-Commitment, and whether they have an equally good explanation as expressivists.

one could take the ethical facts to *just be* facts about our commitments. And, again, this position would afford a straightforward explanation of Supervenience. Views on which facts about our commitments *ground* or *explain* ethical facts might similarly have promising explanations of Supervenience. But, as we discussed earlier, expressivists (we think correctly) have wanted to emphasize the sharp distinction between their accounts and speaker-subjectivism. On pain of collapsing this distinction, it can be no part of expressivism as such, conciliatory or not, that ethical facts are identical to or grounded in facts about our commitments. Against that background, embracing such explanations requires abandoning the distinctive feature of the expressivist view.[68]

### 4.2. "Substantive" ethical explanations of the relevant correlations?

Recall from §1 that conciliatory expressivism is an "ethically neutral" thesis about certain parts of our thought and talk. By contrast, quasi-realism proper is a collection of substantive ethical commitments which are often associated with realism, interpreted within the conciliatory expressivist framework.

Within the conciliatory expressivist framework, one might find the following thought tempting. Object-level claims like Supervenience are *substantive*, like claims about attitude-independence. Hence, they are to be explained by appeal to (a quasi-realist's) substantive ethical commitments.

To see how such an explanation might go, suppose (for simplicity) that you accept a specific substantive ethical commitment: utilitarianism. You might reason as follows:

1. Utilitarianism is true.
2. If utilitarianism is true, then every ethical difference is explained by a difference in facts about the distribution of happiness.

---

[68] Note that one other option would be for the conciliatory expressivist to try to get some of the explanatory benefits of speaker-subjectivism, with respect to the supervenience and reliability challenges, by adopting certain auxiliary commitments beyond their conciliatory expressivism as such, in order to try to "mimic" speaker-subjectivism without actually collapsing into it. For example, they might embrace a form of "quasi-anti-realism", along the lines we discussed earlier in this paper, on which the correctness conditions on one's ethical thoughts are given by idealizing one's psychology. It is beyond the scope of this paper to evaluate this option at length. But there are three main points worth emphasizing on this option. First, most contemporary conciliatory expressivists will very much *not* want to take this option, given that they want to embrace a form of quasi-*realism* instead (and, moreover, that doing that is often a key part of their motivation for adopting conciliatory expressivism in the first place). Second, it is far from clear how much explanatory power any form of "quasi-anti-realism" actually has with respect to the supervenience and reliability challenges. (To illustrate reasons for skepticism here, consider that the most straightforward forms of "quasi-anti-realism" will essentially involve positing a "substantive" ethical explanation of the relevant correlations. Our discussion in §4.2 of "substantive" ethical explanations, and reasons for doubting their explanatory power with respect to the supervenience challenge, is thus directly relevant here). Third, if quasi-anti-realism *does* have explanatory power here, in virtue of the relevant "substantive" ethical commitments that are a part of it, then all the work is being done by ethical commitments that other non-expressivists could adopt too. And so there would be no distinctive explanatory advantage that conciliatory expressivism as such brings to the table. (See here our arguments in §4.2 for more on this point).

3. Facts about the distribution of happiness are natural facts.
4. So, utilitarianism explains why Supervenience is true (from 1-3).
5. So I can explain Supervenience (from 4).

Call this *The Utilitarian Reasoning*. There is nothing special about utilitarianism in this reasoning: parallel reasoning is available to anyone who accepts a consistent and complete ethical theory.

Whether The Utilitarian Reasoning provides a good explanation of Supervenience is controversial.[69] But our core concern here isn't whether The Utilitarian Reasoning is good. Rather, we want to emphasize that *expressivism* plays no role in this reasoning. If it works, it works for anyone who accepts a complete ethical theory. For example, nothing bars the non-naturalist realist utilitarian from arguing in exactly the same way. Our question is whether the expressivist has a dialectical *advantage* with respect to the explanatory challenges. Appealing to a (purported) explanatory resource that anyone can adopt is a failure in this context.

One might argue, however, that The Utilitarian Reasoning is (at least) better *when combined with* certain further expressivist commitments. Here, we consider three strategies: an appeal to further metaphysical commitments, an appeal to minimalism, and an appeal to an expressivist analysis of a crucial piece of vocabulary.

### 4.2.1 property individuation

We are sympathetic to the worry that the "utilitarian explanation" just "moves the bump in the rug".[70] To see this worry, consider:

**Supervenience$_H$** If two metaphysically possible entities (e.g. actions, states of affairs, persons…) are perfectly similar with respect to their relation to the promotion of *happiness*, they are alike in all ethical respects.

Now, Supervenience$_H$ is far from uncontroversial. But our imagined expressivist utilitarian accepts it. And, once accepted, it poses just the same kind of explanatory demand as Supervenience. For happiness and (e.g.) rightness appear to be very different properties. So, we can ask what explains why they covary across all metaphysical possibilities.

Now, as we emphasized in §2, there are metaphysical hypotheses that, if true, would provide the relevant explanation. For example, if, contrary to initial appearances, the property of rightness turned out to be *identical* to the property of happiness maximization, this would explain (what initially seemed like) "covariance" between these "two" properties. After all, there is no mystery about why a property "covaries" with itself.

[69] For discussion, see (Dreier 2019).
[70] For further discussion, see (McPherson 2019b).

Gibbard embraces an "intensional" criterion of property individuation, according to which any "two" properties with exactly similar patterns of instantiation across all metaphysically possible worlds are identical.[71] So, on Gibbard's account, our imagined expressivist utilitarian (for example) is committed to the sort of identity that would explain Supervenience. If we grant the above assumptions, Gibbard is better able to explain Supervenience than the sort of metaethical realist who denies the relevant property identities.

However, expressivism contributes nothing to this explanation. All of the work is done by the combination of normative commitment and metaphysical thesis that philosophers with other, non-expressivist, metaethical views could accept. For example, both of these positions could be accepted by a reductive utilitarian ethical realist.[72]

### 4.2.2 appeal to minimalism

As we pointed out in §1, the conciliatory expressivist has two distinctive types of semantic resources: minimalist and expressivist analyses of certain relevant vocabulary, respectively. Consider in turn whether these resources can improve The Utilitarian Reasoning.

The conciliatory expressivist might propose that their appealing to The Utilitarian Reasoning is better than their metaethical rivals doing so, because the conciliatory expressivist *accepts minimalism*. They might argue that, *given minimalism*, this reasoning is sound, but that it is unsatisfactory if we reject minimalism. We want to make two brief points. First, it is unclear how precisely to spell out the idea that minimalism helps here. For, as we argued in §3, minimalism doesn't undercut the relevant challenges. Second, even if it were to help, it does not show that the conciliatory *expressivist* has a dialectical advantage over the realist. For if minimalism about the semantics of (e.g.) 'fact' is credible, the expressivist's metaethical rivals can appeal to it as well. Again, expressivism does no work in the crucial explanation.

### 4.2.3 appeal to expressivist analyses: explanation

Now let's consider whether an *expressivist* analysis of some relevant vocabulary could help. We focus specifically on *explanatory* vocabulary, as this is arguably the most promising case, given that we are seeking to transform an explanatory challenge. So: would an expressivist account of "explanation" talk help make more promising The Utilitarian Reasoning offered above?

---

[71] (Gibbard 2003, 99).

[72] The expressivist might have a dialectical advantage, if accepting expressivism makes accepting this metaphysical view about properties *more plausible*. We are skeptical that this is so. But if it is, the conciliatory expressivist's dialectical advantages with respect to explaining Supervenience would turn out to be for much more complicated and subtle dialectical reasons than most have thought.

The key question here is whether there is an expressivist account of 'explanation' talk that satisfies two desiderata:

1. It helps improve normative explanations of Supervenience (like The Utilitarian Reasoning),
   and
2. It is credible (which includes that it is a good fit with our felicity judgments about explanatory vocabulary).

It's hard to jointly satisfy these two desiderata. To illustrate, consider a leading contemporary proposal, due to Selim Berker, for what expressivists should say regarding thought and talk about "normative explanation". (Berker ultimately rejects this proposal, but thinks it might be the best that his expressivist opponents can do).

Berker's proposal is that conciliatory expressivists should understand seemingly metaphysically explanatory talk in ethics as expressing psychological "basing" relations among expressed psychological states.[73] Roughly, basing relations hold between beliefs and the conditions that (rationally, at least usually) sustain them. For example, we might say that Charlotte's belief that the fridge is broken is based on her smelling the kitchen. So, on Berker's account, "Kicking dogs is wrong because it causes them pain" expresses *disapproval of kicking dogs that is psychologically based in a belief that kicking dogs causes them pain*.

This account might look like good news for The Utilitarian Reasoning. For example, it does seem like a utilitarian's acceptance of Supervenience can be psychologically based in her accepting utilitarianism, following that reasoning.

Unfortunately, there is a general problem with Berker's account, which is closely related to the contrast we emphasized in §4.1 between explaining a thesis and explaining our rational commitment to that thesis. The core problem is that the basing relation in a person's psychology, and what they take to be the explanatory relation, can be the reverse of each other. For example, Eva's accepting utilitarianism might be *based* on her taking it to be the best explanation of specific ethical facts E1…En. That is, the basing relation might go from cases to principle. But Eva might reasonably believe that utilitarianism explains why facts E1…En hold. That is, her explanatory judgment might go from principle to cases. This sort of idea should be felicitously expressible: it should be conceptually coherent to think that the order of ethical explanation can come apart from the order of epistemological basing in ethics.

Crucially, the *very feature* that appeared to make Berker's account promising as a way to vindicate the utilitarian explanation of supervenience is what makes it implausible as an account of our ordinary explanatory talk. This tension reinforces

[73] (Berker 2020). For a closely connected proposal, which draws on Berker, see (Dreier Manuscript).

a now-familiar general lesson: in general, expressivist accounts can address the explanatory challenges only at the cost of becoming less adequately conciliatory.[74]

This section has argued that the conciliatory expressivist lacks distinctive resources to answer the explanatory challenges. The crucial point was made in the first subsection: because of the distinctive nature of expressivism, the expressivist cannot appeal to certain aspects of their theory of ethical thought in seeking to explain patterns in the distribution of ethical properties. The expressivist can offer explanations (like normative explanations) that do not appeal to these distinctive features, but these explanations are not distinctively available, or favorable, to the expressivist. Finally, we have considered attempts to show these explanations can be made more credible for the conciliatory expressivist given their minimalist or expressivist analyses of certain vocabulary. The appeal to minimalism is unpromising. It is conceivable that an explanation could be developed using the expressivist analyses, but the case of 'explanation' illustrates how difficult this task will be.

## 5. Conclusion

This paper considered two sorts of challenges for metaethical theories. A *supervenience challenge*, to explain why ethical properties cannot vary independently of natural properties, and a *reliability challenge* to explain the reliability of (some of) our ethical beliefs. We claimed that there are only two ways to address explanatory challenges like these: to *undercut* them, or to *answer* them.

We argued that the expressive power that expressivists need in order to be conciliatory appears to bar their undercutting these challenges. More modestly: conciliatory expressivists cannot undercut the challenges in the straightforward way that non-conciliatory expressivists can.

We also argued that expressivists lack distinctive resources for answering the challenges. We have noted two characteristic kinds of failures here. First, accounts that appeal to distinctively expressivist resources tend to explain the wrong thing, for the purposes of the relevant challenge at hand. In short, rather than explain the holding of a relation, they explain (rational) commitment to that relation. Second, the conciliatory expressivist's most promising attempts to explain these relations appeal to explanantia that are wholly independent of conciliatory expressivism

[74] We lack space to carefully consider Derek Baker's alternative account of normative "explanation" talk in (Baker 2021), which he argues can be given a promising "quasi-realist" interpretation. We think that this proposal is in some ways more promising for the quasi-realist than Berker's proposal, but in others less so. In any case, the key dialectical point for us here is that it is equally inapt for the purposes at hand. Notably, if applied in this context (which it wasn't designed for), it appears to suggest that the demand for an explanation of Supervenience is infelicitous, which again suggests a dubious fit with our felicity judgments. (Toppinen 2018, 40) offers a simpler proposal on behalf of the expressivist, according to which to ask why x is right *just is* to ask why to do it. We take Berker's and Baker's work to be an advance on this simplistic proposal.

itself. This independence entails that these explanantia are available for parallel use by many of the expressivist's realist rivals.

We take our argument to go beyond prior challenges to the conciliatory expressivist's ability to undercut or answer these explanatory challenges, by helpfully revealing the general structure of the problem the conciliatory expressivist faces. This feature of our argument suggests pessimism is warranted about the possibility that an excellent novel expressivist solution may be forthcoming. In short, the link between conciliation and expressive power is a substantial barrier to any attempt to undercut the challenges. And the particular kind of gap that the conciliatory expressivist posits between ethical thought and talk on the one hand, and ethical reality on the other, poses a substantial barrier to any attempt to appeal to expressivism (or distinctively expressivist resources) to answer the explanatory challenges.

Our argument *at least* suggests that expressivists cannot claim any advantage over familiar realist views, with regard to the supervenience and reliability challenges. And, strikingly, some realists can offer satisfying answers to these challenges that seem unavailable to conciliatory expressivists. Given that these are widely considered to be among the deepest and most difficult challenges to the realist in metaphysics and epistemology respectively, this casts doubt on the relative plausibility of expressivism, especially compared to the latter sorts of realist views. Of course, this is compatible with the possibility that some variant of conciliatory expressivism is, overall, the most plausible metaethical theory. But if we are right, this will not be in part because of a straightforward expressivist dialectical advantage with respect to these explanatory challenges. It will, instead, be because of some other reason(s): for example, perhaps because conciliatory expressivism does a particularly good job at smoothly explaining relevant links between ethical judgments and motivation while also successfully capturing various linguistic patterns.

Beyond providing strong reason to accept our conclusion, our argument helps clarify what anyone hoping to undercut or answer these challenges on behalf of conciliatory expressivism needs to do. For example: anyone wanting to undercut the challenges needs to provide a clear story about how their undercutting is compatible with an adequately conciliatory story about the expressive power of our ethical language, and with platitudes about the rational significance of evidence of unreliability. Similarly, anyone wanting to answer the challenges needs to explain object-level phenomena, not simply patterns in our commitments. And they need to explain how their account is compatible with the expressivist's commitment to avoid positing that conclusions about thought and talk entail substantive ethical conclusions.

We have not considered every possible way to address these challenges. For example, it is conceivable that a novel expressivist analysis of some key piece of apparently metaphysical vocabulary might enable the conciliatory expressivist to undercut or answer the challenges. And, as we noted in §1.2, minimalist and expressivist accounts do not exhaust the linguistic resources that the expressivist

might use to achieve conciliation. The conciliatory expressivist may wish to investigate a variety of other ways of "deflating" apparently metaphysical vocabulary.

Our discussion does, however, suggest general reason for pessimism about such strategies as ways of addressing these challenges. As we illustrated in our discussion of Camil Golub's undercutting strategy, and of Selim Berker on 'explanation' talk, expressivist analyses that promise natural undercutting or explanatory accounts tend to do so only by appealing to resources that make them fail to be conciliatory in striking ways.

The punchline, then, is this: it is possible that a specific version of conciliatory expressivism could be developed that would turn out to have an advantage over many realist views in addressing these challenges. But if so, it will be because of some distinctive and highly non-obvious aspect of this specific view, and *not* simply because it is a form of expressivism. Moreover, the difficulties we've canvassed in this paper provide general reasons to be skeptical that such a view will emerge – or at least while also delivering a version of metaethical expressivism that makes good on the core motivations for the view in the first place, and which avoids collapsing into a rival view that expressivists have standardly gone to great lengths to distinguish their view from.

In closing, we sketch four ways that our arguments and conclusions are significant beyond the specific "conciliatory expressivist versus realist" dialectic in metaethics that has framed this paper. These concern, respectively, the dialectic with non-conciliatory forms of expressivism, the prospects for "hybrid" expressivist views meeting the challenge, relevance for expressivist hypotheses outside of metaethics, and relevance for expressivism understood as a "conceptual ethics" view.

First, consider the dialectic between conciliatory and non-conciliatory versions of expressivism. Non-conciliatory expressivists need to reject the felicity of much everyday ethical discourse. This is often taken to be a decisive reason to favor the conciliatory version of the view. But at least some versions of non-conciliatory expressivists have a different advantage: they may be able to undercut one (or both) of the supervenience and reliability challenges. As we emphasized earlier, this is not so for all forms of non-conciliatory expressivism, with respect to both challenges. But, still, insofar as some such non-conciliatory views can make good on this promise, it may complicate the case for the common contemporary view that conciliatory versions of expressivism are "the only game in town" for metaethical expressivists.[75]

Second, consider so-called "hybrid" expressivist accounts – on which, roughly, ethical thoughts are claimed to have a both a "cognitive" and "non-cognitive" component.[76] There are initial reasons to believe that our argument can be extended to target most such views. This is because "hybrid" expressivisms characteristically

[75] Compare (McPherson 2022).
[76] For an overview of hybrid expressivism, see (Toppinen 2017).

share two key features with the conciliatory view we have discussed, and these features drive our argument. First, hybrid views characteristically provide the expressive power needed to felicitously state the challenges. Second, they characteristically ban entailments from ethical thought to "ethical reality", so as to avoid collapsing into sophisticated forms of speaker-subjectivism.

Third, we take our discussion to likely generalize beyond ethics to conciliatory expressivist accounts in other areas (e.g., epistemology, philosophy of law, and metaphysics).[77] This is because the central structure of our argument did not appeal to anything special about the *ethical* character of metaethical expressivism, but rather focused on the sorts of resources available to a conciliatory expressivist about *any* topic.

Finally, this paper focused on conciliatory expressivism as a descriptive hypothesis about our *actual* ethical thought and talk. However, the apparent inability of conciliatory expressivism to address these explanatory challenges is also a potential reason to reject conciliatory expressivism as a normative proposal in "conceptual ethics", regarding how to reform our ethical concepts, or about which broadly "ethical-ish" concepts we should use.[78] In short, that's because it's plausible that, other things being equal, we should want to engage in a form of "ethical-ish" thought and talk which would allow us to address these sorts of challenges in a satisfactory way.

## Acknowledgments

For helpful comments and discussion, thanks to Farbod Akhlaghi, Derek Baker, Chloe Banino, Selim Berker, David Braddon-Mitchell, Matthew Chrisman, Jamie Dreier, Andy Egan, David Enoch, David Faraci, Stephen Finaly, Allan Gibbard, Nadeem Hussain, Zöe Johnson King, Frank Jackson, Sebastian Köhler, Niko Kolodny, Robert Kraut, Catherine Lapey, Rishi Joshi, Barry Maguire, Emily Massey, Peter Railton, Adrian Russian, Geoff Sayre-McCord, Joshua Schechter, Sharon Street, Amie Thomasson, Christine Tiefensee, Patrick Todd, Kevin Toh, Marcus Willaschek, and Seth Yalcin. Earlier versions of this paper were presented at Bowling Green University, Dartmouth College, the Frankfurt School of Finance and Management, the University of Sydney, and the Chapel Hill Normativity Workshop. Thanks to everyone who participated in those conversations.

## Works Cited

Ayer, A.J. 1936/1971. *Language, Truth, and Logic*. New York: Penguin Books.

[77] For some examples of expressivist (or at least expressivist-*like*) views in areas outside of metaethics, see e.g., (Price 2011), (Thomasson 2020), (Toh 2011), (Yalcin 2007), (Pérez Carballo 2014), and (Chrisman 2022).

[78] For further discussion about how we think about "conceptual ethics" projects in this area, see (McPherson and Plunkett 2024) and (McPherson and Plunkett 2025a).

Baker, Derek. 2021. If You're Quasi-Explaining, You're Quasi-Losing. In *Oxford Studies in Metaethics Volume 16*, edited by R. Shafer-Landau: Oxford University Press.

Bedke, Matthew. 2017. Cognitivism and Non-Cognitivism. In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett. New York City Routledge.292-307.

Berker, Selim. 2020. Quasi-Dependence. In *Oxford Studies in Metaethics, Vol. 15*, edited by R. Shafer-Landau. Oxford: Oxford University Press.195–218.

Björnsson, Gunnar, and Tristram McPherson. 2014. Moral Attitudes for Non-Cognitivists: Solving the Specification Problem. *Mind* 123 (489):1-38.

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

———. 1998. *Ruling Passions*. Oxford: Clarendon Press.

———. Manuscript. Sharon Street on the Independent Normative Truth as Such.

Camp, Elisabeth. 2017. Metaethical Expressivism. In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett. New York City: Routledge.87-101.

Chrisman, Matthew. 2005. Review of Allan Gibbard's *Thinking How to Live*. *Ethics* 115 (2):406-412.

———. 2022. *Belief, Agency, and Knowledge: Essays on Epistemic Normativity*: Oxford University Press.

Cuneo, Terence. 2008. Moral Realism, QuasiRealism, and Skepticism. In *The Oxford handbook of skepticism*, edited by J. Greco: Oxford University Press.176.

———. 2013. Properties for Nothing, Facts for Free? Expressivism's Deflationary Gambit. *Oxford Studies in Metaethics* 8:223-51.

———. 2017. Quasi-realism. In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett. New York: Routledge.626-642.

———. 2020. Can expressivism have it all? *Philosophical Studies* 177 (1):219-241.

Dreier, James. 1996. Expressivist embeddings and minimalist truth. *Philosophical Studies* 83 (1):29-51.

———. 2004. Meta-Ethics and the Problem of Creeping Minimalism. *Philosophical Perspectives* 18:23-44.

———. 2010. When do goals explain the rules that advance them? *Oxford Studies in Metaethics* 5.

———. 2012. Quasi-Realism and the Problem of Unexplained Coincidence. *Analytic Philosophy* 53 (3):269-287.

———. 2015. Explaining the Quasi-Real. In *Oxford Studies in Metaethics, Vol 10*. Oxford: Oxford University Press.273-297.

———. 2019. Is there a supervenience problem for robust moral realism? *Philosophical Studies* 176 (6):1391-1408.

———. Manuscript. Is There a Quasi-Queer 'Because'?

Egan, Andy. 2007. Quasi-realism and fundamental moral error. *Australasian Journal of Philosophy* 85 (2):205 – 219.

Enoch, David. 2010. The epistemological challenge to metanormative realism: how best to understand it, and how to cope with it. *Philosophical Studies* 148 (3):413-438.

———. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

Faraci, David. 2017. On Leaving Room for Doubt: Using Frege–Geach to Illuminate Expressivism's Problem with Objectivity. *Oxford Studies in Metaethics* 12:244-264.

Faraci, David, and Tristram McPherson. 2017. Ethical Judgment and Motivation. In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett. New York: Routledge.308-323.

Finlay, Stephen. Manuscript. Alice the Camel Has No Humps: Why we should only be quasi-expressivists.

Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.

———. 2011. How Much Realism? Evolved Thinkers and Moral Concepts In *Oxford Studies in Metaethics*, edited by R. Shafer-Landau. Oxford: Oxford University Press.33-51.

Golub, Camil. 2017a. Expressivism and Realist Explanations. *Philosophical Studies* 174 (6):1385-1409.

———. 2017b. Expressivism and the Reliability Challenge. *Ethical Theory and Moral Practice* 20 (4):797-811.

Hare, R.M. 1952/1991. *The Language of Morals*. Oxford: Oxford University Press.

Jackson, Frank, and Philip Pettit. 1998. A Problem for Expressivism. *Analysis* 58 (4):239-251.

Klagge, James C. 1988. Supervenience: Ontological and ascriptive. *Australasian Journal of Philosophy* 66 (4):461-70.

Köhler, Sebastian. 2021. How to Have Your Quasi-Cake and Quasi-Eat it Too. *The Canadian Journal of Philosophy* 51:204-220.

Kraut, Robert. 1993. Robust deflationism. *Philosophical Review* 102 (2):247-263.

McPherson, Tristram. 2012. Ethical Non-naturalism and the Metaphysics of Supervenience. In *Oxford Studies in Metaethics, Vol. 7*, edited by R. Shafer-Landau. Oxford: Oxford University Press.205-234.

———. 2019a. Evaluative predicates as classificatory devices? *Philosophical Studies* 176 (6):1439-1451.

———. 2019b. Supervenience in Ethics In *The Stanford Encyclopedia of Philosophy (Winter 2019 Edition)*, edited by E. N. Zalta. URL = <https://plato.stanford.edu/archives/win2019/entries/supervenience-ethics/>.

———. 2022. Expressivism Without Minimalism. In *Meaning, Decision, and Norms: Themes From the Work of Allan Gibbard*, edited by B. Dunaway and D. Plunkett. Ann Arbor, MI: Maize Books, University of Michigan Publishing.147-170.

McPherson, Tristram, and David Plunkett. 2017. The Nature and Explanatory Ambitions of Metaethics. In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett. New York: Routledge.1-25.

———. 2024. Metaethics and the Conceptual Ethics of Normativity. *Inquiry* 67 (1):93-126.

———. 2025a. After Metaethics. *Philosophers' Imprint* 23 (23):1-17.

———. 2025b. The Supervenience of the Normative and the Autonomy of Essence: Lessons from Leary's Hybrid Gambit. In *The Future of Normativity*, edited by S. Kirchin. Oxford: Oxford University Press.187-213.

Olson, Jonas. 2017. Error Theory in Metaethics. In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett. New York City: Routledge.58-71.

Pérez Carballo, Alejandro. 2014. Structuring Logical Space. *Philosophy and Phenomenological Research* 92 (2):460-491.

Plunkett, David. 2020. Conceptual truths, evolution, and reliability about authoritative normativity. *Jurisprudence* 11 (2):169-212.

Price, Huw. 2011. *Naturalism Without Mirrors*. New York: Oxford University Press.

———. 2015. From Quasirealism to Global Expressivism – and Back Again? In *Passions and Projections: Themes from the Philosophy of Simon Blackburn*, edited by R. N. Johnson and M. Smith. Oxford: Oxford University Press.134–152.

Rosen, Gideon. 2020. What is Normative Necessity? In *Metaphysics, Meaning and Modality: Themes from Kit Fine*, edited by M. Dumitru. Oxford: Oxford University Press.205-233.

Schechter, Joshua. 2010. The Reliability Challenge and the Epistemology of Logic. *Philosophical Perspectives* 24 (1):437-464.

———. 2017. Explanatory Challenges in Metaethics. In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett. New York: Routledge.443-458.

———. Manuscript. Does Expressivism Enjoy an Epistemological Advantage Over Realism?

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

———. 2008b. Expression for Expressivists. *Philosophy and Phenomenological Research* 76 (1):86-116.

———. 2010. *Noncognitivism in Ethics*. New York City: Routledge.

———. 2014. Does expressivism have subjectivist consequences? *Philosophical Perspectives* 28 (1):278-290.

Street, Sharon. 2006. A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127 (1):109-166.

———. 2011. Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have It Both Ways. In *Oxford Studies in Metaethics, vol. 6*, edited by R. Shafer-Landau. Oxford: Oxford University Press.

Sturgeon, Nicholas L. 2009. Doubts about the Supervenience of the Evaluative. In *Oxford Studies in Metaethics Vol. 4*, edited by R. Shafer-Landau. Oxford: Oxford University Press.53-92.

Thomasson, Amie L. 2020. *Norms and Necessity*. New York: Oxford University Press.

Toh, Kevin. 2011. Legal Judgments as Plural Acceptances of Norms. In *Oxford Studies in Philosophy of Law: Volume 1*, edited by L. Green and B. Leiter. Oxford: Oxford University Press.107-137.

Toppinen, Teemu. 2017. Hybrid Accounts of Ethical Thought and Talk. In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett. New York City Routledge.243-259.

———. 2018. Non-Naturalism Gone Quasi: Explaining the Necessary Connections between the Natural and the Normative. *Oxford Studies in Metaethics* 13:25-47.

Venesmaa, Vilma. 2021. Explaining our Knowledge of Normative Supervenience. *Oxford Studies in Metaethics* 16:233-256.

Woods, Jack. 2017. The Frege-Geach Problem. In *The Routledge Handbook of Metaethics*, edited by T. McPherson and D. Plunkett. New York: Routledge.226-242.

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

———. 2018. Expressivism by force. In *New Work on Speech Acts*, edited by D. Fogal, D. W. Harris and M. Moss. Oxford: Oxford University Press.400-430.