Fallibility for Expressivists

Bob Beddor

Abstract

Quasi-realists face the challenge of providing a plausible analysis of acknowledgments of moral fallibility (e.g., *I believe that lying is wrong, but I might be mistaken*). This paper develops a new analysis of these acknowledgments, according to which they express moral uncertainty. After advertising the advantages of this analysis, I take up the question of how to understand moral uncertainty in expressivist terms.

1   The Challenge

According to moral expressivism, the function of moral discourse is not to describe the world. Rather, it's to express desire-like attitudes. The main challenge for expressivism comes from the fact that moral discourse behaves much like descriptive discourse. Expressivists in the quasi-realist tradition seek to meet this challenge: they aim to reconcile expressivism with the realist trappings of moral language.

Can this goal be achieved? In recent years, some authors have claimed that quasi-realists are unable to account for *acknowledgments of moral fallibility* (Egan 2007; Köhler 2015). An example: Ava has just completed a term paper arguing that eating meat is morally wrong. She's confident that her thesis is true. Still, she admits that it's possible that her thesis is mistaken. Ethics is hard, and she's aware that many smart people disagree with her. She is thus inclined to say things like:

(1)   a. I believe eating meat is wrong.
      b. But I might be mistaken.[1]

How should we understand these acknowledgments, if moral beliefs are just desire-like states? Call this 'The Fallibility Challenge.'

---

[1]Beddor and Goldstein (2018) call these conjunctions 'concessive belief attributions' based on their resemblance to concessive knowledge attributions (sentences of the form, ⌜*I know* ϕ, *but I might be mistaken*⌝). Interestingly, while such CKAs are typically judged infelicitous, their belief-involving counterparts sound perfectly natural. For discussion of CKAs, see, a.o., Lewis (1996); Stanley (2005); Hetherington (2013).

2   The Path Ahead

Quasi-realists have not left this challenge unanswered. The most detailed quasi-realist account of acknowledgments of moral fallibility is due to Blackburn. According to Blackburn, to acknowledge your moral fallibility is to express doubts about whether you would retain your moral beliefs under idealized conditions—that is, conditions of full information, careful reflection, and so forth.[2] However, this 'Idealization Analysis' faces serious difficulties. Egan (2007) and Köhler (2015) argue that it cannot make sense of the full range of acknowledgments of moral fallibility. And Schroeder (2013) objects that it looks *ad hoc* and implausible from a compositional perspective.

In light of these difficulties, this paper develops a new solution to the Fallibility Challenge. I begin with the observation that we typically acknowledge our moral fallibility using epistemic modals (e.g., *might*). Over the last decade, there has been a surge of interest in expressivist treatments of epistemic modals. According to 'credal expressivism,' the function of epistemic modals is not to describe the world. Rather, it's to express agents' credences.[3] While credal expressivism and moral expressivism are usually developed in isolation from one another, they make for natural partners. And if we integrate them, we get a promising new analysis of discourses like (1). On the resulting analysis, acknowledgments of moral fallibility are *expressions of moral uncertainty*. I argue that this 'Credal Analysis' neatly avoids the problems facing the Idealization Analysis.

An obvious objection is that my solution assumes that expressivists can make sense of moral uncertainty. But can they? The second half of this paper responds to this worry by developing an expressivist theory of moral credences. Drawing on ideas from Gibbard (1990, 2003), Sepielli (2012), and Goldstein (2016), I propose that an agent's moral uncertainty consists in the degree to which they plan to adopt various reactive attitudes. If this account is on the right track, we'll have solved two problems for quasi-realists at one fell swoop: we'll have shown how they can make sense of both moral uncertainty and moral fallibility.

3   Idealization and its Discontents

3.1   The Idealization Analysis

Let's begin by taking a closer look at the standard quasi-realist response to the Fallibility Challenge. Blackburn articulates the idea thus:

> How can I make sense of my own fears of fallibility? Well, there are a number of things that I admire: for instance, information, sensitivity, maturity, imagination, coherence. I know that other people show defects in these respects,

---

[2] See Blackburn (1973, 1998), as well as Horgan and Timmons (2015); Lam (forthcoming). See also Ridge (2015) for a variant of this analysis, cashed out in a hybrid expressivist framework.

[3] This idea traces back to Ramsey (1926). For contemporary developments, see Yalcin (2007, 2011); Rothschild (2012); Moss (2013).

and that these defects lead to bad opinions. But can I exempt myself from the same possibility? Of course not… So I can think that perhaps some of my opinions are due to defects of information, sensitivity, maturity, and imagination, and coherence. (1998: 318)

Let's see how this applies to our example involving Ava. Expressivists claim that in uttering (1a) (*I believe that eating meat is wrong*) Ava reports a fact about her conative attitudes—for example, that she disapproves of eating meat. The Idealization Analysis maintains that by following up with (1b) (*But I might be mistaken*), Ava acknowledges that this conative attitude might stem from some defect. An improved sensibility might not share her conative attitude.

3.2   First Difficulty: Idealized Sensibilities Might Err

Despite its appeal, the Idealization Analysis faces two serious difficulties. First, Egan (2007) observes that it has trouble accommodating the thought that even *idealized* sensibilities might err. To unpack this worry, imagine that Ava ruminates as follows:

It seems someone could know all of the non-moral facts, while still failing to know all of the moral facts. Of course, this failure could be due to further defects on their part—say, lack of imagination or sensitivity. But need it be? This isn't obvious: just because someone is fully informed, sensitive, and imaginative, it's not clear that she'll be guaranteed to know all of the moral truths.

Following this line of thought, Ava judges:

(2)   Even if my belief that eating meat is wrong survives idealization, this belief might be mistaken.

The Idealization Analysis predicts that (2) expresses an incoherent judgment: it amounts to judging that even if her belief survives idealization, it might not survive idealization. This is an uncomfortable result: (2) seems perfectly coherent.[4]

Proponents of the Idealization Analysis might push back. Maybe once we get clear on the relevant sense of 'idealization,' we'll recognize that (2) is incoherent after all. Blackburn takes this line in reply to Egan. According to Blackburn, if a belief is false, '[T]hen an improvement is clearly on the cards, namely, replacing it with the truth' (2009: 206). So any belief that survives idealization is, by definition, guaranteed to be true.

However, this response comes with a serious cost. If quasi-realists pursue this line, they forfeit any claim to have provided a non-circular analysis of acknowledgments of

---

[4]My statement of this worry departs in a number of respects from Egan's formulation, in part so as to sidestep features of Egan's formulation that Blackburn (2009) deems objectionable. In particular, Blackburn insists that the notion of 'idealization' is normative, and hence should be understood in expressivist fashion. My formulation of the challenge is consistent with this point.

moral fallibility. On the view that results, an acknowledgment that a moral judgment J might be in error is analyzed as an acknowledgment that J might not survive idealization. But 'idealization' is itself analyzed in terms of the avoidance of moral error.[5]

3.3 Second Difficulty: Semantic Plausibility

Further trouble is in store. Note that Ava's utterance of (1) contains an epistemic possibility modal (*might*). Given this observation, a natural strategy for solving the the Fallibility Problem would be to start by getting clear on the semantic contribution of possibility modals, and use this to derive an analysis of (1).

But the Idealization Analysis does not pursue this strategy. Instead, it proceeds in a 'piecemeal, *ad hoc* fashion' (Schroeder 2013: 416). Consequently, it generates implausible predictions concerning how embedding a sentence under *might* affects its meaning. As Schroeder notes, the Idealization Analysis holds that the following two sentences are equivalent:

(3)   It might not be wrong to eat meat.

(4)   I might not disapprove of eating meat if I were an idealized agent.

However, expressivists typically insist that the following are *not* equivalent:

(5)   It is not wrong to eat meat.

(6)   I would not disapprove of eating meat if I were an idealized agent.

After all, expressivists maintain that there is a crucial difference between expressing a mental state *m* and reporting that one is in *m*. According to expressivists, (5) expresses lack of disapproval of eating meat; it does not report that one would not disapprove of eating meat under such-and-such conditions. But then why would embedding (5) and (6) under the same epistemic modal (*might*) generate equivalent sentences?

4   The Credal Analysis

In order to develop a more satisfactory solution, I propose proceeding in two steps. First, give a general analysis of the meaning of *might*. Here I focus on an analysis that should be particularly welcome to moral expressivists: *credal expressivism.* Second, use this analysis to derive an analysis of acknowledgments of moral fallibility.

---

[5]Some may suggest that expressivists shouldn't try to offer a non-circular analysis of acknowledgments of moral fallibility. (Cf. Blackburn 2009.) However, as Köhler (2015) and Ridge (2015) emphasize, one goal of the quasi-realist program is to explain the mental states we express when we make various realist-sounding claims. So quasi-realists need to give some story about what mental state Ava expresses when she utters (1b). Without some analysis of acknowledgments of moral fallibility, it's unclear how the quasi-realist can fulfill this commitment.

4.1 Credal Expressivism

Credal expressivism is typically advanced in opposition to a 'descriptivist' semantics for epistemic modals, according to which utterances of epistemic modals purport to describe facts about the world. Consider:

(7) It might be raining.

According to descriptivists, (7) describes the world as being a particular way—for example, as being one where the prejacent (*It's raining*) is consistent with the information available to some contextually determined agents.[6]

Credal expressivists deny this. According to credal expressivists, (7) does not have descriptive truth conditions at all. Instead, it expresses the speaker's positive credence in the possibility that it's raining.

Moral expressivism and credal expressivism make for natural allies. For starters, the two views are structurally analogous. Both claim that certain bits of language are not used to describe the world, but rather to express certain mental states. Furthermore, both views are typically motivated by similar arguments.

First, there's what I'll call 'the subject matter argument.' Credal expressivists often object that descriptivism offers an implausible picture of the subject matter of modal beliefs. Having a modal belief, according to this objection, does not require having beliefs *about* anyone's information. Yalcin gives the example of Fido, who believes his owner might give him a bone (2011: 308). Does this require that Fido believes that the receipt of a bone is compatible with his information (or anyone else's, for that matter)? This seems like a stretch. More plausibly, it only requires that Fido assigns some credence to the possibility that his owner will give him a bone.[7]

In a similar vein, moral expressivists often object that rival accounts provide an implausible picture of the subject matter of moral beliefs. Take, for example, a simple subjectivist view, according to which S believes that stealing is wrong iff S believes that S disapproves of stealing. Against this, expressivists often complain that the *reasons* for believing that stealing is wrong are importantly different from the reasons for believing that you disapprove of stealing. The latter concern your psychology; they could be provided by a psychoanalyst or a brain scan. By contrast, the former sort of reasons are moral: they are reasons to disapprove of stealing in the first place.

Second, there's what I'll call 'the argument from disagreement.' Building on Price (1983), Yalcin (2011) argues that if I disagree with your claim that it might be raining, I do not thereby disagree with the claim that *your* information is compatible with the possibility that it will rain. Here too, parallel arguments have been offered for moral expressivism. Against simple subjectivism, it is often objected that if I disagree with your claim that stealing is wrong, I do not thereby disagree with the claim that *you* disapprove

---

[6]See, a.o., Kratzer (1981); Dowell (2011).
[7]See also Rothschild (2012); Moss (2013).

of stealing.[8]

The upshot: those who find the arguments for moral expressivism convincing should also be drawn to credal expressivism.

4.2   Solving the Fallibility Challenge

On to the second step of the solution. Credal expressivism delivers a general analysis of acknowledgments of fallibility. Take any instance of the schema:

(8)   a.   I believe ϕ.

   b.   But I might be mistaken.

According to what I'll call the 'Credal Analysis,' in asserting (8a), the speaker asserts that she believes ϕ. But in following up with (8b), she hedges by expressing her positive credence in $\neg\phi$, and hence her lack of certainty in ϕ. This is coherent, given the plausible assumption that belief does not require certainty.[9]

I propose this holds for both moral and non-moral instances of ϕ. Thus when Ava utters (1a), she reports her belief that eating meat is wrong. And in following up with (1b), she hedges this by expressing her lack of certainty that eating meat is wrong.

An immediate worry for the Credal Analysis is that talk of moral *credences* seems incompatible with moral expressivism. I will devote the second half of this paper to addressing this worry. For now, let us bracket this concern and see how the Credal Analysis avoids the difficulties facing the Idealization Analysis.

The first problem for the Idealization Analysis was that one can coherently worry that idealized sensibilities might err, as expressed in Ava's utterance of (2) (*Even if my belief that eating meat is wrong survives idealization, this belief might be mistaken*). The Credal Analysis has no trouble here. According to the Credal Analysis, (2) expresses Ava's conditional uncertainty. Specifically, it expresses the state of assigning some positive credence to the prospect that eating meat isn't wrong, conditional on her belief that eating meat is wrong surviving idealization. Since this is a coherent credal state, the Credal Analysis correctly predicts that (2) is coherent.

Some might wonder whether the problem re-emerges in a different form. Does the Credal Analysis predict that it would be incoherent for Ava to worry that an ideally

---

[8]See Stevenson (1937); Schroeder (2010): 69-70.

[9]What is the relation between belief and credence? For our purposes, we need not take a stand. One could embrace a 'Lockean' view, according to which beliefs are just credences that exceed some threshold. Alternatively, one could allow that belief entails high credence but deny that belief is reducible to credence; perhaps belief is a *sui generis* state.

A further question concerns what sort of mental state is expressed by assertions of atomic sentences (e.g., *eating meat is wrong*). The simplest answer would be that asserting ϕ expresses a belief in ϕ. However, given that belief does not require certainty, this raises the question of why 'epistemic contradictions' of the form ⌜ϕ *and might* $\neg\phi$⌝ sound incoherent, unlike (8). In response to this problem, one option is to hold that assertions express a state stronger than belief, such as certainty. Another option is to retain the idea that assertions express beliefs, but adopt an expressivist analysis on which the conjunction ⌜ϕ *and might* $\neg\phi$⌝ is semantically inconsistent. (See Beddor and Goldstein (2018) for relevant discussion.)

rational agent's moral *certainties* might be mistaken? If so, isn't this prediction equally problematic?

In assessing this objection, we need to take care: much depends on how we formulate the relevant worry. One way is via a conditional such as:

(9)   Even if I were rationally certain that eating meat is wrong, I might still be mistaken.

The Credal Analysis makes sense of this worry. (9) expresses Ava's positive credence that eating meat isn't wrong, conditional on her ideally rational counterfactual self becoming certain that eating meat is wrong. And this is a coherent credal state.

Alternatively, Ava might worry that she is (currently) rationally certain of some particular claim and that this claim might be mistaken:

(10)   a.   I'm rationally certain that eating meat is wrong.
       b.   ? But I might be mistaken.

However, *this* sounds odd. The Credal Analysis explains why. If (10a) were true, Ava could not coherently assign some credence to the possibility that eating meat is not wrong. And so she could not coherently be in the mental state expressed by (10b).

On to the second problem for the Idealization Analysis: it looks implausible and *ad hoc* from a compositional perspective. Here too the Credal Analysis is in the clear. Take any possibility modal, combine it with some sentence ϕ, and you'll get a sentence that expresses the speaker's positive credence in ϕ. Consequently the Credal Analysis avoids predicting that (3) (*It might not be wrong to eat meat*) is equivalent to (4) (*I might not disapprove of eating meat if I were an idealized agent*). Indeed, (3) is not equivalent to any claim about what agents might disapprove of under such-and-such conditions.

5   Credences for Expressivists: Two Challenges

The Credal Analysis gives rise to an obvious worry. I've spoken freely about expressing moral *credences*. But isn't this a cognitivist notion? If expressivism is true, how can we make sense of having, say, a .8 credence that eating meat is wrong?

A preliminary point: quasi-realists were already committed to providing an account of moral credence, independent of the Fallibility Challenge. After all, quasi-realists aim to make sense of ordinary moral discourse. And ordinary discourse is rife with talk of moral uncertainty, e.g.:

(11)   Ava isn't sure whether she did the right thing.

(12)   Darren is confident that lying is wrong, but he's even more confident that murder is wrong.

To analyze these claims, quasi-realists need to develop a theory of moral uncertainty. So in using moral credences to solve the Fallibility Problem, expressivists do not incur any *new* explanatory burdens.

Still, the question remains: how should expressivists understand moral credence? An initially tempting strategy is to identify moral belief with some conative state that comes in degrees, such as desire or disapproval. Expressivists could then hold that moral credences correspond to degrees of this state. However, this simple strategy faces two fatal problems.

The first problem, due to Smith (2002), is that this simple account does not have enough *structure* to model moral judgment. Smith points out that moral judgment is gradable along two different dimensions. The first dimension—'certitude'—is your degree of confidence that, say, eating meat is wrong. The second—'importance'—is *how* wrong you take eating meat to be. These dimensions can vary independently: you could have a middling credence that eating meat is very wrong, or a high credence that it is slightly wrong. Any account that reduces moral credence to degrees of desire or disapproval will be unable to draw this distinction. Call this 'Smith's Challenge.'

A further challenge arises from the fact that moral credence has a lot in common with descriptive credence. To give one particularly clear example, both are governed by the same coherence constraints. It's commonly held that rationality requires our credences to conform to the probability axioms—a requirement that applies to both moral and descriptive credences. By contrast, it's dubious that degrees of desire or disapproval are subject to this requirement. Indeed, it's unclear that we can meaningfully talk about a .8 degree of desire or disapproval.[10] Call this the 'Commonality Challenge.'

In what follows, I'll sketch an expressivist account of moral credence that has a better shot of overcoming these challenges. §5 tackles Smith's challenge; §6 addresses the Commonality Challenge.[11]

6   Solving Smith's Challenge

According to a prominent tradition, there is a close connection between *wrongness*-judgments and 'reactive attitudes' (Strawson 1974) such as blame and resentment. This idea has its roots in Mill (1861). According to Mill, to claim that someone acted wrongly is to say that he ought to be punished, 'if not by law, by the opinion of his fellow-creatures; if not by opinion, by the reproaches of his own conscience.'[12] Gibbard (1990) gives Mill's idea an expressivist twist. On Gibbard's view, to judge that eating meat is wrong is to adopt some conative attitude towards a syndrome of reactive attitudes towards meat-eating—a syndrome that might include resentment and outrage towards carnivores, as well as admiration for vegetarians. Call this, 'the Reactive Attitude Account.'

---

[10]See Bykvist and Olson (2012); Staffel (forthcoming).

[11]While I aim to show that there is a plausible way of developing an expressivist account of moral credence, nothing in the Credal Analysis hinges on the details of this account. Those who prefer some alternative expressivist treatment of moral credences can still help themselves to the Credal Analysis in order to solve the Fallibility Problem. For examples of alternative treatments, see Eriksson and Olinder (2016); Ridge (forthcoming).

[12]Mill (1861): chp. V, par. 4.

Sepielli (2012) shows that this approach has the resources to solve Smith's Challenge. On the Reactive Attitudes Account, wrongness judgments involve two components: (i) a cluster of reactive attitudes, (ii) a conative attitude towards this cluster. As Sepielli notes, we can use these two components to model Smith's two dimensions of moral judgment. Variations in importance consist in variations in the severity of the reactive attitudes. Variations in certitude consist in variations in the strength of the conative attitude towards these reactive attitudes.

To illustrate, consider a simple (too simple, for reasons to be discussed shortly) version of the Reactive Attitudes Account, according to which the relevant conative attitude is *approval* and the relevant reactive attitude is *blame*. Then having a high credence that eating meat is slightly wrong amounts to having a high degree of approval for mildly blaming carnivores. And having a moderate credence that eating meat is very wrong amounts to having a moderate degree of approval for severely blaming carnivores.

Some might worry that the extra complexity of the Reactive Attitudes Account is *ad hoc.* But it is independently motivated, since it helps solve two further problems for expressivists. First up: the 'Negation Problem' (Unwin 1999; Schroeder 2008). This problem stems from the observation that there are three ways we can stick negation into a moral belief report, each of which ascribes a different mental state. But we can't model these differences using the simple expressivist view that a *wrongness*-judgment is just some conative attitude (say, disapproval) directed towards an action:

not believing eating meat is wrong	not disapproving of eating meat
believing not eating meat is wrong	disapproving of not eating meat
believing eating meat is not wrong	?

As Schroeder (2008) and Sepielli (2012) observe, the additional structure in the Reactive Attitudes Account overcomes this difficulty[13]:

not believing eating meat is wrong	not approving of blaming for eating meat
believing not eating meat is wrong	approving of blaming for not eating meat
believing eating meat is not wrong	approving of not blaming for eating meat

Next up: the 'Moral Attitude Problem' (Miller 2003). This problem arises from the fact that not every conative attitude makes for a *moral* belief. I can desire a cookie without thinking it would be wrong for me to go cookieless. The challenge for expressivists is to say what sort of conative attitude makes for a distinctly moral judgment. The Reactive Attitudes Account provides an answer. My desire for a cookie does not make for a *wrongness*-judgment because I do not approve of blaming or shaming myself in the event that I fail to consume one.

---

[13]Though see Baker (2018) for critical discussion.

7 Solving the Commonality Challenge

7.1 Plan Expressivism

While the Reactive Attitudes Account solves Smith's Challenge, it leaves the Commonality Challenge unanswered. We've said that moral certitude consists in degrees of some conative attitude directed towards various reactive attitudes. But why does this conative attitude have so much in common with descriptive credence?

In response, we should start by noting that a version of this challenge already faces every expressivist account of *outright* moral belief. After all, moral and descriptive beliefs have a lot in common (Schroeder 2010: 96-97). To give just one example, they are subject to the same coherence constraints: it is incoherent to believe both p and $\neg p$, regardless of whether p is descriptive or moral. These similarities are *prima facie* puzzling if moral belief is a conative attitude, since it seems coherent to have contradictory desires.

In order to solve the Commonality Challenge for outright belief, one promising strategy is to identify moral belief with an attitude that is desire-like in some respects but belief-like in others.[14] This is why Gibbard (2003) identifies moral beliefs with *plans* (i.e., intentions). Unlike many conative attitudes, plans display many of the normative and functional hallmarks of belief.

Start with the normative. Just as it is incoherent to have contradictory beliefs, so it is incoherent to have contradictory plans. If I plan to spend the afternoon at the library, it would be incoherent for me to simultaneously plan to stay home all day.[15]

On to the functional. If I believe p, then I will be disposed to avoid reconsidering the question of whether p. Similarly, Bratman (1987: 18-19) observes that if I plan to spend all afternoon at the library, then I will be disposed to avoid reconsidering whether to stay at the library. (By contrast, if I merely desire to spend the afternoon at the library, I might find myself debating whether to go the movies instead.) More generally, the *plan expressivist* strategy is to explain the similarities between descriptive and moral beliefs in terms of the similarities between beliefs and plans.

Does this strategy extend to degrees of belief? Just like belief, planning come in degrees. We are fully committed to executing some of our plans; we are less than fully committed to executing others. A natural suggestion is to identify moral certitude with *degrees of plans* directed towards various reactive attitudes. We could then use the similarities between degrees of plans and degrees of belief in order to account for the similarities between moral and descriptive credence.

---

[14]For a different approach to the Commonality Challenge, see Beddor (2019).

[15]Here's another normative similarity. Reasons for belief can be *weighed* against one another: a reason to believe p can be outweighed by a stronger reason to believe $\neg p$. Maguire (2018) argues that the reasons for adopting certain affective attitudes don't work like this. If a suffering loved one dies, the reasons for being distressed do not seem to be outweighed by the reasons in favor of feeling relief. (Indeed, Maguire takes this to cast doubt on the very idea that there are reasons for affective attitudes.) Here too, planning seems to pattern with belief. The reasons in favoring of planning to go library-ward can be outweighed by the reasons in favor of planning to stay home.

In order to implement this strategy, we would need to give a theory of degrees of planning, one that predicts that degrees of planning really do resemble descriptive credences in the same ways that moral credences do. The rest of this section considers two theories that hold promise.

7.2 Plans as Beliefs

According to *cognitivism about plans*, plans are just beliefs.[16] On a simple version of this view, to plan to ϕ is to believe that one will ϕ, or that one is disposed to ϕ.

This simple view requires qualification. Suppose a student is told by an oracle that they will fail an upcoming exam. The student might believe the oracle without thereby planning to fail. In light of such cases, most cognitivists maintain that plans are a special sort of belief—beliefs that figure into practical reasoning in a particular way. For example, one might hold that to plan to ϕ is to believe that one will ϕ, and for this belief to dispose one to try to ϕ.[17] This explains why our student does not plan to fail the exam. While they believe that they will fail, this belief does not dispose them to try to fail.

If plan expressivists embrace cognitivism about plans, they'll have a simple story about why moral belief has so much in common with descriptive belief. On the resulting view, to believe that eating meat is wrong is to have a special sort of descriptive belief. It's to believe that one will (or is disposed to) have various reactive attitudes towards meat-eating, and for this belief to dispose one to try to adopt these reactive attitudes. No wonder moral and descriptive belief have so much in common!

This solution extends straightforwardly to credences. According to a natural generalization of the cognitivist view, S plans to degree n to ϕ just in case they have credence n that they will (or are disposed to) ϕ, and this credence plays an action-guiding role in practical reasoning. Combine this with plan expressivism, and we get the view that moral credences are just a special type of descriptive credence. It thus comes as no surprise that the two states share so many properties.

Some might worry that cognitivism about planning is incompatible with plan expressivism. Isn't the whole idea behind expressivism that moral beliefs are *noncognitive* states? But I think this worry can be defused. On the resulting view, moral beliefs are *not* analyzed as beliefs with some distinctly moral content. Rather, they are analyzed as beliefs about the reactive attitudes one will adopt. Thus the view makes good on the expressivist commitment to explaining moral thought and talk in entirely non-moral, naturalistic terms. And since plans are understood as beliefs that play an action-guiding role in practical reasoning, this view also vindicates the expressivist idea that moral beliefs play a distinctive role in motivating action.

---

[16]See e.g., Velleman (1989); Marušić and Schwenkler (2018).

[17]See Marušić and Schwenkler (2018) for a somewhat different way of developing this idea.

7.3 Plans as Multi-Track Dispositional States

Cognitivism about plans is controversial. Let us therefore consider the leading noncognitivist theory of plans, due to Bratman (1987). On Bratman's view, planning is a *sui generis* attitude whose functional role can be characterized in terms of various dispositions. In particular, planning to ϕ involves (i) being disposed to ϕ, (ii) being disposed to seek out the means of ϕing, (iii) being disposed to avoid reconsidering whether to ϕ. If plan expressivists were to embrace this view, would they still be able to solve the Commonality Challenge?

On Bratman's view, there are important differences between plans and beliefs. But there are also important similarities. For example, the third dispositional feature of plans is shared by beliefs: as noted earlier, someone who believes p is disposed to avoid reconsidering whether p. Moreover, Bratman argues that his view underwrites the idea that plans, unlike desires, are subject to many of the same normative requirements as beliefs—in particular, a consistency requirement (1987: 31). If there are enough similarities between Bratmanian plans and beliefs, perhaps we can use these to solve the Commonality Challenge for outright belief.

Can Bratman's view of plans also solve the Commonality Challenge for credences? I think there are grounds for optimism. Specifically, expressivists can appeal to an independent line of argument, developed in Goldstein (2016), that a Bratmanian view of plans supports the thesis that degrees of planning are subject to probabilistic coherence constraints. If this is right, then while we might not have yet explained *all* of the commonalities between moral credence and descriptive credence, we'll have made considerable progress: we'll have shown why they are governed by the same coherence requirements.

First things first: how should we understand degrees of planning, on a Bratmanian view? As Goldstein (2016) observes, dispositions are gradable. A statue can be more or less fragile; a material can be more or less flammable. The dispositions that constitute planning are no exception. This suggests a natural extension of Bratman's view: the degree to which an agent plans to ϕ is determined by the strengths of the various dispositions that constitute planning to ϕ.

Next, how should we understand dispositions? In the literature on dispositions, one promising approach is to analyze dispositions in terms of quantification over some domain of cases C in which some stimulus condition obtains (Manley and Wasserman 2008; Vetter 2014). A major selling point of these modal accounts is that they shed light on the gradability of dispositions. On a modal account, the degree to which x has some disposition D is determined by the proportion of C-cases in which x manifests D. For example, suppose we want to analyze the degree to which a statue is fragile. The C-cases will be circumstances in which the statue is dropped from various heights, or struck with varying degrees of force. The higher the proportion of such cases in which the statue breaks, the more fragile the vase.

Suppose we combine a dispositional account of plans with a modal account of dispositions. Then the degree to which one plans to ϕ will correspond to the proportion

of C-cases in which one manifests the various dispositions that constitute planning to ϕ. Given this model, Goldstein argues, a case can be made for thinking that a rational agent's degrees of planning should satisfy the probability axioms. While the argument is complex, here is a broad-brushstrokes overview.

Start with the Non-Negativity axiom, which in this context means that for any ϕ, the degree to which one plans to ϕ is never less than 0. This is trivially validated by the current model: the lowest possible degree of planning is to manifest the relevant dispositions in 0% of the C-cases.

Turn next to Normalization, which says that an agent should maximally plan to satisfy the tautologous property ($\top$). On the present model, this amounts to manifesting the dispositions that constitute planning to $\top$ in all C-cases. On a Bratmanian view of plans, the relevant dispositions will be: (i) a disposition to satisfy $\top$, (ii) a disposition to search for the means necessary to do so, (iii) a disposition to avoid reconsidering whether to do so. As Goldstein observes, a case can be made for thinking every rational agent will manifest these dispositions in every C-case, for any plausible candidate for 'C'. After all, it is metaphysically necessary that every agent satisfies $\top$. Thus, every agent will trivially have the means to satisfy $\top$. Moreover, it would be pointless to reconsider whether to satisfy $\top$.

Finally, consider Additivity, according to which the degree to which an agent plans to satisfy a disjunctive property $\vee$ is the sum of the degrees to which she plans to satisfy $\vee$'s disjuncts. Now on the view under consideration, the degree to which she plans to satisfy $\vee$ is reflected in the proportion of C-cases in which she manifests the dispositions that constitute this plan. Goldstein argues, again using a Bratmanian view of plans, that the proportion of C-cases in which a rational agent manifests these dispositions will be the sum of the cases in which she manifests the dispositions that constitute planning $\vee$'s disjuncts.

Obviously, there is much one can debate here. But if an argument along these lines succeeds, we will have shown that a Bratmanian view of plans can be used to explain why degrees of planning are subject to the same coherence requirements as descriptive credence.[18]

7.4   Taking Stock

Which of these two theories of degrees of planning should the expressivist prefer?

This depends on how much similarity expressivists want to secure between moral and descriptive uncertainty. We've seen that both approaches can explain some important similarities between the two states. However, on the dispositional theory moral uncertainty is still a different sort of state than descriptive uncertainty; they just happen

---

[18]Goldstein also provides an independent argument that degrees of planning are subject to probabilistic requirements. This argument involves extending accuracy dominance arguments for probabilism to degrees of intentions. In the interest of space, I defer discussion of this argument to another occasion. For related discussion, see Staffel (forthcoming).

to share certain structural and normative features.

Some expressivists might embrace this result. After all, many expressivists maintain that moral and descriptive beliefs differ in their motivational profiles. So perhaps these expressivists will welcome the consequence that moral and descriptive uncertainty have different functional roles. For those who find this unsatisfying, the cognitivist theory will prove more appealing. After all, the cognitivist theory says that moral uncertainty is just a type of descriptive uncertainty.

8   Synthesis and Conclusion

This paper developed a new analysis of acknowledgments of moral fallibility. According to the Credal Analysis, acknowledgments of moral fallibility are expressions of moral uncertainty. I went on to supplement this analysis with an expressivist account of moral uncertainty. On the resulting view, your credence that ϕing is wrong is the degree to which you plan to adopt various reactive attitudes towards ϕing.

To tie it all together, return to Ava's utterance of (1) (*I believe eating meat is wrong. But I might be mistaken*). The first conjunct reports that Ava has a particular moral belief. On the view developed here, for her to have this belief is for her to plan to engage in a cluster of reactive attitudes towards meat-eating. But in uttering the second conjunct, she expresses that she isn't certain that eating meat is wrong—i.e., that she assigns some credence to the possibility that it is not wrong. This means that she doesn't *maximally* plan to adopt the relevant reactive attitudes towards meat-eating.

How we understand this state of mind depends on which theory of degrees of planning we adopt. On the cognitivist theory, for Ava to plan to degree .8 to blame carnivores is for her to have a .8 credence that she will do so. On the dispositionalist theory, it's for her to manifest the various plan-constituting dispositions in 80% of some relevant class of worlds. Either way, we'll have cashed out moral credences in expressivist-friendly terms.

By way of conclusion, let me address three potential objections.

Objection: Your view has a lot of moving parts. Aren't we being asked to stomach a bunch of *ad hoc* solutions to various problems?

Reply: While I agree my account is complex, I think we can assuage the worry that it is *ad hoc*. First, note that this paper has sought to accomplish two different tasks on the expressivist agenda: explaining moral fallibility and explaining moral uncertainty. Expressivists were already committed to accomplishing the latter task, aside from any consideration of the former. (After all, expressivists owe us some story about our everyday ascriptions of moral uncertainty, e.g., (11)-(12).) So my account of moral uncertainty is not an *ad hoc* fix to the Credal Analysis. Rather, it's something that expressivists independently needed.

Second, the hallmark of an *ad hoc* solution is that there is no independent motivation for it. But each piece of my account is independently motivated. Start with the Credal

Analysis. The primary argument for the Credal Analysis was that it avoided the two problems facing the Fallibility Analysis. However, the idea that epistemic modals express speakers' credences was motivated by independent arguments (the subject matter argument, the argument from disagreement)—arguments that parallel standard motivations for moral expressivism.

Next, turn to the account of moral uncertainty. This had two parts: the Reactive Attitudes Account and plan expressivism. The Reactive Attitudes Account was primarily motivated by its ability to solve Smith's Challenge. However, it was independently supported by its ability to solve two further problems for expressivists: the Negation Problem and the Moral Attitude Problem. The plan expressivist component was primarily motivated by its ability to solve the Commonality Challenge. This solution relied on the premise that degrees of planning have much in common with degrees of belief—a premise that was independently motivated by consideration of two promising theories of degrees of planning.

Objection: On the account of moral credence defended here, it should be incoherent to claim:

(13) I (fully) plan to blame meat eaters, but I'm not sure that eating meat is wrong.

But such utterances seem coherent.[19]

Reply: In response to this challenge, let me raise three points. First, there is a question about the data. While I agree that (13) doesn't sound like an overt contradiction, it also doesn't strike me as obviously coherent. If someone asserted (13), I suspect most people would be puzzled about the speaker's attitudes vis-à-vis eating meat.

Second, anyone who accepts that wrongness judgments are tied to reactive attitudes will have reason to question the coherence of (13). Assume that to judge that ϕing is wrong is to judge that ϕing warrants blame. Then any reason for being less than certain that eating meat is wrong is a reason to refrain from blaming carnivores. Next, assume that any reason against ϕing is also a reason against planning to ϕ. Conclusion: any reason for being less than certain that eating meat is wrong is also a reason against planning to blame meat eaters.

The third point is more concessive. Let's grant that (13) is coherent. Then while this would be a problem for my account, it would also be a problem for *any* expressivist account of moral belief. On any expressivist account, moral beliefs are conative attitudes. But now pick your favorite conative attitude and plug it into (13), e.g.:

(14) I (fully) disapprove of eating meat, but I'm not sure that eating meat is wrong.

You'll get a sentence that sounds just as coherent as (13), if not more so.[20]

---

[19]Thanks to a referee for raising this point.

[20]Cf. Woods (2014), who questions whether expressivists can explain the coherence of: *I believe eating meat is wrong, but I don't disapprove of it.*

For the purposes of this paper, I won't take a stand on how expressivists should repair their account to predict the coherence of these sentences (if, indeed, they should). Any repair that works for (14) should also work for (13).

Objection: You haven't explained what sort of mental state someone is in when they worry that idealized sensibilities might err. You said that when Ava utters (2), she expresses a state of *conditional* moral uncertainty. But you haven't told us how to make sense of this state.

Reply: We often make conditional plans (or 'contingency plans'). Currently I do not plan to buy groceries, since I don't know whether the store is open. But, conditional on the store being open, I plan to buy them. A natural proposal is to identify conditional moral uncertainty with degrees of conditional plans. Thus when Ava asserts (2), she is expressing *conditional indecision*. Even conditional on her belief that eating meat is wrong surviving idealization, she is less than fully committed to blaming carnivores.

Some might worry that this is an odd state of mind for Ava to be in. In ordinary cases of conditional indecision, there is some further information that would help us settle on a course of action. By contrast, Ava is in a state of *fundamental conditional indecision*—a sort of indecision that, by her own lights, no further information could resolve.

But expressivists can dispel the apparent oddity of this state by pointing to other examples of fundamental conditional indecision. Suppose Ava is trying to decide on a career: surgeon or philosopher? On one way of filling out the case, even if Ava were fully informed of all the descriptive facts about these two career paths, she would still feel torn. I propose that it is this state of mind that expressivists should use when trying to understand Ava's worry that an idealized sensibility might err.[21]

References

Derek Baker. Expression and Guidance in Schroeder's Expressivist Semantics. *Erkenntnis*, 83:
    829–852, 2018.

Bob Beddor. Noncognitivism and Epistemic Evaluations. *Philosophers' Imprint*, 19(10):1–27, 2019.

Bob Beddor and Simon Goldstein. Believing Epistemic Contradictions. *Review of Symbolic Logic*,
    11:87–114, 2018.

Simon Blackburn. Moral Realism. In Casey, editor, *Morality and Moral Reasoning*. Methuen,
    London, 1973.

Simon Blackburn. *Ruling Passions*. Clarendon Press, Oxford, 1998.

Simon Blackburn. Truth and *A Priori* Possibility: Egan's Charge Against Quasi-Realism. *Aus-
    tralasian Journal of Philosophy*, 87(2):201–213, 2009.

Michael Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge,
    MA, 1987.

Krister Bykvist and Jonas Olson. Against the *Being For* Account of Normative Certitude. *Journal of Ethics and Social Philosophy*, 6(2):1–8, 2012.

J.L. Dowell. A Flexible Contextualist Account of Epistemic Modals. *Philosophers' Imprint*, 11(14): 1–25, 2011.

Andy Egan. Quasi-Realism and Fundamental Moral Error. *Australasian Journal of Philosophy*, 85 (2):205–219, 2007.

John Eriksson and Ragnar Francén Olinder. Non-Cognitivism and the Classification Account of Moral Uncertainty. *Australasian Journal of Philosophy*, 94(4):719–735, 2016.

Allan Gibbard. *Wise Choices, Apt Feelings.* Harvard University Press, Cambridge, MA, 1990.

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

Simon Goldstein. A Preface Paradox for Intention. *Philosophers' Imprint*, 16(14):1–20, 2016.

Stephen Hetherington. Concessive Knowledge-Attributions: Fallibilism and Gradualism. *Synthese*, 190(14):2835–2851, 2013.

Terry Horgan and Mark Timmons. Modest Quasi-Realism and the Problem of Deep Moral Error. In Johnson and Smith, editors, *Passions and Projections: Themes from the Philosophy of Simon Blackburn.* Oxford University Press, Oxford, 2015.

Sebastian Köhler. What is the Problem with Fundamental Moral Error? *Australasian Journal of Philosophy*, 93(1):161–165, 2015.

Angelika Kratzer. The Notional Category of Modality. In Eikmeyer and Rieser, editors, *Words, Worlds, and Contexts: New Approaches in Word Semantics.* W. de Gruyter, Berlin, 1981.

Garrett Lam. Making Quasi-Realists Admit of Fundamental Moral Fallibility. *Australasian Journal of Philosophy*, forthcoming.

David Lewis. Elusive Knowledge. *Australasian Journal of Philosophy*, 74(4):549–567, 1996.

Barry Maguire. There are No Reasons for Affective Attitudes. *Mind*, 127(507):779–805, 2018.

David Manley and Ryan Wasserman. On Linking Dispositions and Conditionals. *Mind*, 117(465): 59–84, 2008.

Berislav Marušić and John Schwenkler. Intending is Believing: A Defense of Strong Cognitivism. *Analytic Philosophy*, 59(3):309–340, 2018.

J.S. Mill. *Utilitarianism.* 1861.

Alexander Miller. *Introduction to Contemporary Metaethics.* Polity, Cambridge, 2003.

Sarah Moss. Epistemology Formalized. *Philosophical Review*, 122(1):1–43, 2013.

Huw Price. Does 'Probably' Modify Sense? *Australasian Journal of Philosophy*, 61(4):396–408, 1983.

Frank Ramsey. Truth and Probability. In Mellor, editor, *Philosophical Papers.* Cambridge University Press, Cambridge, 1926.

Michael Ridge. I Might be Fundamentally Mistaken. *Journal of Ethics and Social Philosophy*, 9(3), 2015.

Michael Ridge. Normative Certitude for Expressivists. *Synthese*, forthcoming.

Daniel Rothschild. Expressing Credences. *Proceedings of the Aristotelian Society*, 112(1):99–114, 2012.

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

Mark Schroeder. *Noncognitivism in Ethics.* Routledge, New York, 2010.

Mark Schroeder. Two Roles for Propositions. *Noûs*, 47(3):409–430, 2013.

Andrew Sepielli. Normative Uncertainty for Non-Cognitivists. *Philosophical Studies*, 160:191–207, 2012.

Michael Smith. Evaluation, Uncertainty, and Motivation. *Ethical Theory and Moral Practice*, 5(3): 305–320, 2002.

Julia Staffel. Expressivism, Normative Uncertainty, and Arguments for Probabilism. In Szabó Gendler and Hawthorne, editors, *Oxford Studies in Epistemology*, volume 6. Oxford University Press, forthcoming.

Jason Stanley. Fallibilism and Concessive Knowledge Attributions. *Analysis*, 65(2):126–131, 2005.

Charles L. Stevenson. The Emotive Meaning of Ethical Terms. *Mind*, 46:14–31, 1937.

P.F. Strawson. Freedom and Resentment. In *Freedom and Resentment and other Essays*, pages 1–25. Methuen, London, 1974.

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

David Velleman. *Practical Reflection*. CSLI Publications, 1989.

Barbara Vetter. Dispositions Without Conditionals. *Mind*, 123(489):129–156, 2014.

Jack Woods. Expressivism and Moore's Paradox. *Philosophers' Imprint*, 14:1–12, 2014.

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

Seth Yalcin. Nonfactualism about Epistemic Modality. In Egan and Weatherson, editors, *Epistemic Modality*. Oxford University Press, Oxford, 2011.