A Way Out for Normative Fallibilism

Introduction

Epistemologists often make normative claims of the form ‘it is permissible (or impermissible) to have attitude A in situation S’. For example, ‘it is permissible (or impermissible) to believe P if you were told that P by a clairvoyant that is known to have a good track record’. However, what attitude we should have in a given situation is often far from obvious to us, and normative claims in epistemology continue to be the topic of much philosophical debate.

It seems that mistaken beliefs about what attitude we should have in a given situation can themselves be epistemically permissible. If a respected professor offers me a convincing argument that adopting attitude A is *impermissible* in situation S even though attitude A is actually *permissible* in situation S, it seems like it could nonetheless be epistemically permissible for me to believe that attitude A is impermissible in S on the basis of their testimony. Similarly, if a respected professor offers me a convincing argument that adopting attitude A is *permissible* in situation S even though attitude A is *impermissible* in situation S, it seems like it could be epistemically permissible for me to believe that attitude A is permissible in S on the basis of their testimony.

Let us use ‘fallibilism’ refer to the view that we can have permissible false beliefs about normative claims in epistemology. In ‘Rationality’s Fixed Point (or: In Defense of Right Reason)’, Michael Titelbaum (2015) formulates a challenge to fallibilism. He argues that the view is inconsistent with a plausible anti-akrasia principle in epistemology. This anti-akrasia principle says that it is not permissible for you to believe P while also believing that it is impermissible for you to believe P. Titelbaum calls this challenge to fallibilism the No Way Out argument.

This paper is structured as follows: in section 1 I will outline fallibilism and Titelbaum’s No Way Out argument. In section 2 I will show that the No Way Out argument extends to fallibilists who embrace a distinction between ideal and non-ideal epistemic permissibility. In section 3 I introduce the ‘Top Down’ view, and argue that this view can offer the fallibilist a viable way out of the No Way Out argument. In section 4 I motivate the Top Down view using ‘level-connecting’ principles. And in section 5, I respond to Titelbaum’s second challenge to fallibilism: the Self-Undermining argument by arguing that it involves a problematic form of self-reference.

1 The No Way Out Argument

Epistemic normative fallibilists hold that agents can have false beliefs[1] about which doxastic attitudes are permissible or impermissible in a given situation, and that these false beliefs can be epistemically permissible in some important sense. Someone who holds that we can have false beliefs in normative claims, but that these false beliefs are never epistemically permissible, is not a fallibilist in the sense described here.

In order to formulate the fallibilist's view clearly, let A be a doxastic attitude in a proposition (belief, disbelief, suspension, etc.) and $\neg A$ be a lack of that attitude.[2] Let $O_S(A)$ mean that attitude A is obligatory in S, and let $P_S(A)$ mean that attitude A is permissible in S. The fallibilist that I refer to in this paper is one that endorses something at least as strong as the following:

> **Weak Normative Fallibilism (WNF)**: There exists a situation S in which attitude A is epistemically permissible (impermissible) in S, and it is permissible in S to believe that attitude A is epistemically impermissible (permissible) in S
> ($\exists S \exists A(P_S(A) \wedge P_S(B(\neg P_S(A))))$ and $\exists S \exists A(\neg P_S(A) \wedge P_S(B(P_S(A)))))$

Here 'permissible' and 'impermissible' refer solely to epistemic permissibility. And 'situation' refers to whatever features of the agent's environment or mental state (for example, her evidence) are taken to be relevant to the permissibility of her attitudes.[3]

Suppose I'm in a situation S. Can I permissibly believe 'it is impermissible in S to believe things on the basis of perceptual evidence', even if it is not actually impermissible to believe things on the basis of perceptual evidence in S? It seems so. For example, suppose that a friend that I know to be highly reliable told me that it's impermissible to believe things on the basis of perceptual evidence in situation S. Or suppose that I happen to have taken something that I believe to be LSD but was nothing more than inert blotting paper. In cases like these, it seems like I can be permitted to believe that it is impermissible for me to believe things on the basis of perceptual evidence, even if this is not actually true. We could also defend Weak Normative Fallibilism by appealing to ought implies can principles, or by pointing to difficulties that arise for normative infallibilists. However, for the purpose of this paper is simply to defend fallibilism against a recent challenge. I will therefore refrain from giving a more robust defense of Weak Normative Fallibilism, beyond pointing to its prima facie plausibility in cases like these.

Our beliefs about whether an attitude is epistemically permissible or not also seem to play a role in determining what we are permitted to believe. Suppose that I have been born into a particularly dishonest community, and so I believe that it's impermissible to adopt a belief solely on the basis of someone's testimony. Imagine that someone in my community tells me that it will rain tomorrow, and – solely on the basis of their testimony – I come to believe that it will rain

[1] I will set aside any discussion of credal fallibilism in this paper in order to focus on *outright* beliefs, but I believe that arguments considered here apply to both sorts of attitudes.

[2] Note that lacking an attitude of belief need not be the same as disbelief: many think that an agent may lack the attitude of belief without it following that she has the attitude of disbelief. I take no view on this matter here.

[3] By appealing to 'situations' we can avoid misidentifying the No Way Out and Self-Undermining objections as objections for one particular normative theory in epistemology, such as evidentialism or reliabilism.

tomorrow. Suppose I don't revise my belief that it's impermissible to adopt a belief solely on the basis of someone's testimony. Then I will be left believing that it is impermissible for me to believe that it will rain tomorrow on the basis of testimony alone, while also holding the belief - on the basis of testimony alone - that it will rain tomorrow. There seems to be something wrong with this combination of beliefs. An overall belief state that includes both a belief that p and a belief that believing p is impermissible is known as an *epistemically akratic* belief state, and many have argued that it is not epistemically permissible to have a belief state that is epistemically akratic. Those who believe that epistemic akrasia is not epistemically permissible will accept something like the following anti-akrasia principle, suggested by Titelbaum:

> **Anti-Akrasia Principle (AA)**: $\neg\exists S\,\exists A(P_S(A \wedge B(\neg P_S(A))))$
> (There is no situation S in which an agent is jointly permitted to believe that attitude A is impermissible in S and to adopt attitude A in S)

Here I have formulated Anti-Akrasia as a wide scope principle: it is of the form $\neg\exists S\,\exists A(P_S(A \wedge B(\neg P_S(A))))$ instead of the narrow scope form $\neg\exists S\,\exists A(A \wedge P_S(B(\neg P_S(A))))$. The narrow scope version of the principle says that *if* an agent does adopt attitude A – even if it was impermissible for her to do so – then she is not permitted to believe that A is impermissible. The wide scope version of the principle says that an agent is not permitted to jointly adopt attitude A and the belief that A is impermissible. But if adopting attitude A is impermissible then one way to satisfy this wide scope principle is to cease to adopt attitude A. I have also formulated Anti-Akrasia as a *synchronic* rather than a *diachronic* principle of rationality: it is a principle that applies to an agent's attitude state at a single time, rather than a principle about how she should change her beliefs across time.

I believe that cases of epistemically akratic beliefs like the 'it will rain' example given above strike us as intuitively impermissible, which is consistent with the Anti-Akrasia principle. Moreover, principles like Anti-Akrasia have been widely endorsed in the literature as general principles of rationality, even if some people believe that they are not without exceptions.[4] So I will assume here that views which cannot endorse a principle like Anti-Akrasia are, all else being equal, less appealing than views that can endorse a principle like Anti-Akrasia.

In a recent article, Michael Titelbaum[5] argues that Anti-Akrasia is in tension with fallibilism. In what he calls the No Way Out argument, he argues that if one accepts an anti-akrasia principle like Anti-Akrasia, then one must reject fallibilism. The No Way Out argument against fallibilism that Titelbaum gives is as follows (p. 267, 2015):

> 'Begin by supposing... that we have a case in which an agent's situation rationally requires that attitude A, yet also rationally permits an overall state containing the belief that A is rationally forbidden to her. Now consider that permitted overall state, and ask whether A appears in it or not. If the permitted overall state does not contain A, we have a contradiction with our supposition that the agent's situation requires A.

[4] See Horowitz, S. (2014), 'Epistemic Akrasia', *Noûs*, 48: 718–744 for a recent defense of this view.

[5] Titelbaum, Michael G.: 'Rationality's Fixed Point (or: In Defense of Right Reason)', Ch. 9, Tamar Szabó Gendler and John Hawthorne (Eds), *Oxford Studies in Epistemology 5* (2015): 253–294

> (That supposition says that every overall state rationally permissible in the situation contains A). So now suppose that the permitted overall state includes A. Then the state includes both A and the belief that A is forbidden in the current situation. By the Akratic Principle, this state is not rationally permissible, contrary to supposition once more. This completes our reductio.'

Titelbaum is asking us to consider the possible situation in which we have a permissible belief that attitude A is impermissible. We are then asked whether this is a situation in which we can have attitude A. The answer is no: although attitude A is permissible, it is not *jointly* permissible with the belief that attitude A is impermissible. But we said from the start that attitude A was obligatory. If we assume that an attitude that is obligatory is also permissible, which seems highly plausible, then this leads us to a contradiction.

One thing that some people might want to reject in this argument is the claim that some attitudes can be *obligatory* rather than merely permissible. But we can follow Titelbaum in assuming that an attitude is obligatory if and only if the attitude is permissible and there is no alternative attitude available to the agent that is permissible in the agent's situation. This is a rather weak notion of epistemic obligation, and so it may be more palatable to those who are inclined to reject the claim that some attitudes can be obligatory rather than merely permissible. We don't need a stronger notion of epistemic obligation than this in the No Way Out argument.

The kind of fallibilism that Titelbaum presents a challenge for is actually somewhat stronger than the fallibilism expressed by Weak Normative Fallibilism. To see this, it will be helpful to try to give an explicit formulation of the No Way Out argument. Let us assume that an agent knows that she is in situation S.[6] We can then formulate Titelbaum's argument as follows:

> **Strong Normative Fallibilism (SNF)**: $\exists S \exists A(O_S(A) \land P_S(B(\neg P_S(A))))$
> (There exists a situation S in which an attitude A is obligatory, but it is permissible for the agent in S is permitted to believe that A is impermissible in S)
>
> **Anti-Akrasia (AA)**: $\neg\exists S \exists A(P_S(A \land B(\neg P_S(A))))$
> (There is no situation S in which an agent is permitted to believe both that attitude A is impermissible in S and to adopt attitude A in S)
>
> **Deontic Inference Rule (DIR)**: $O_S(A_1) \land P_S(A_2) \rightarrow P_S(A_1 \land A_2)$
> (If one attitude is obligatory in situation S while another attitude is permissible in situation in S, then it is permissible to adopt both attitudes in S)

[6] Titelbaum (p. 262-4, 2015) wants to avoid saying that an agent is obligated to have true beliefs about what is epistemically obligatory in her current given situation even if she has an a posteriori false beliefs about the content of her situation or about what attitudes her current overall attitude state contains. However, it is not clear that this No Way Out argument would not also apply to cases that involve mistakes of this sort.

> It is easy to show that these three premises cannot be jointly satisfied. By NF and DIR, there exists a situation S such that it's jointly permissible to adopt attitude A and to believe that A is not permissible. But this contradicts AA, which says that no such situation exists. Therefore AA and DIR jointly entail that NF is false.[7]

I have included a deontic inference rule in this formulation of the argument. Although this deontic inference rule is not explicit in Titelbaum's argument, it summarizes the move that Titelbaum is making when he formulates his dilemma informally in the argument above.

This argument presents a problem for Strong Normative Fallibilism, since both Anti-Akrasia and the Deontic Inference Rule are highly plausible principles of epistemic permissibility. It seems plausible that – in most cases, at least – an agent is not permitted to adopt epistemically akratic attitudes like 'P and I'm not permitted to believe P'. It also seems plausible that if, in the same situation, one attitude is obligatory and another attitude is permissible, then one can permissibly adopt both attitudes in that situation, as the Deontic Inference Rule states. There are, of course, cases in which an obligatory attitude A_1 might make an otherwise permissible attitude A_2 impermissible, and it would not be permissible to adopt both attitudes in this kind of situation. But that would not be the kind of situation in which the Deontic Inference Rule applies, since that would not be a situation in which A_1 is obligatory and – as a result of this fact – A_2 is *not* permissible but is instead impermissible. As such, the Deontic Inference Rule seem like a highly plausible principle of epistemic permissibility, and we should avoid rejecting it if possible.

Some might respond to this argument by pointing out that akratic attitudes are permissible in at least some situations. Notice, however, that Titelbaum doesn't need something as strong as Anti-Akrasia for the No Way Out argument to create problems for the fallibilist. Suppose we think that anti-akrasia principles are true in some but not all situations. The No Way Out argument would still show that all of the situations in which we are not permitted to be akratic are also situation in which an agent cannot falsely believe that some obligatory attitude is impermissible. And presumably cases in which we are permitted to be epistemically akratic are the exceptions and not the rule. So even a commitment to weaker anti-akrasia principles than Anti-Akrasia may not save Strong Normative Fallibilism in cases where we are not permitted to be akratic.

However, the No Way Out argument does not present a direct challenge to Weak Normative Fallibilism, even if we grant both Anti-Akrasia and the Deontic Inference Rule. Fallibilists who accept Weak normative Fallibilism but not Strong Normative Fallibilism can hold that it is only ever permissible to believe that A is impermissible in cases where A is permissible but not obligatory.[8] So Weak Normative Fallibilism is jointly consistent with Anti-Akrasia and the Inference Rule, even though Strong Normative Fallibilism is not.

[7] Here is an alternative version of the argument: from DIR it follows that $\neg P_S(A \wedge B(\neg P_S(B(A)))) \rightarrow \neg(O_S(A) \wedge P_S(B(\neg P_S(A))))$. So if an agent is permitted in S to believe that A is not permissible in S, then A is not obligatory in S.

[8] Note that this restriction is perfectly compatible with Weak Normative Fallibilism as it is stated above. While attitude A being obligatory entails that attitude A is permissible, it is not the case that attitude A being permissible entails that attitude A is obligatory. So although the view that agents can permissibly believe that an obligatory attitude is impermissible entails Weak Normative Fallibilism, Weak Normative Fallibilism does not entail that agents can permissibly believe that an obligatory attitude is impermissible.

In order to present a challenge to Weak Normative Fallibilism directly, we might try to extend Titelbaum's argument to cases in which an attitude A is permissible but not obligatory. But this extended No Way Out argument would not present any conflict with Anti-Akrasia. Weak Normative Fallibilism says that there is some situation S such that $P_S(A)$ and $P_S(B(\neg P_S(A)))$. But it may be the case that A is only permissible if $B(\neg P(A))$ is not permissible, and vice versa. We might try to claim that $P_S(A) \wedge P_S(B)$ entails $P_S(A \wedge B)$. But this claim is very implausible. An agent may be permitted to believe Q and also be permitted to believe R without being permitted to believe them both at the same time (for example, Q and R might be equally plausible, but mutually inconsistent). If we accept that this is possible, then the Deontic Inference Rule, Weak Normative Fallibilism and Anti-Akrasia can all be true.[9]

But many fallibilists committed to Weak Normative Fallibilism will also want to accept the claim that it can be permissible to believe that A is impermissible in cases where attitude A is obligatory, which Strong Normative Fallibilism commits us to. After all, the same reasons to believe that we can be mistaken about what is permissible – for example, that we can receive misleading inductive or testimonial evidence that a permissible attitude is impermissible – also seem to apply to attitudes that are obligatory rather than merely permissible. And Titelbaum's No Way Out argument shows that fallibilists who accept Strong Normative Fallibilism and want to maintain that it is not permissible to have epistemically akratic beliefs face serious difficulties.

2 Ideal and Non-Ideal Epistemic Permissibility

In the previous section I assumed that there is only one kind of epistemic permissibility. But fallibilists might want to distinguish between different kinds of epistemic permissibility: for example, *ideal* epistemic permissibility and *non-ideal* epistemic permissibility.[10] Let us define an ideally permissible attitude as an attitude that an agent with unlimited cognitive powers but limited evidence would be permitted to have in a given situation. And let us define a non-ideally permissible attitude as an attitude that an agent with limited cognitive powers and limited evidence would be permitted to have in a given situation. Since we can identify impermissibility with 'not permissible to' and obligation with 'not permissible not to', the ideal/non-ideal distinction extends to impermissible and obligatory attitudes as well.

Suppose that the fallibilist holds that these two kinds of epistemic permissibility sometimes give different verdicts about a given attitude. For example, suppose she holds that agents are *ideally* obligated to have true beliefs about which attitudes are ideally permissible or impermissible in their current situation, but they are *non-ideally* permitted to have false beliefs about which attitudes are ideally permissible or impermissible in their current situation. To make this view easier to understand, consider the following example:

[9] In other words, Anti-Akrasia presents a problem for anyone who thinks that it can be *jointly* permissible to have attitude A and believe that attitude A is impermissible.

[10] Here I have used the ideal/non-ideal distinction, but this is not an important choice. We could have chosen objective and subjective epistemic permissibility, or epistemic and instrumental epistemic permissibility. The important thing is just that these accounts adopt different kinds of permissibility.

> Eric has been doing research into different principles of induction. He has encountered some particularly strong arguments for the weak induction principle *W* over the strong induction principle *Z*, and so he comes to believe that he is obligated to adopt attitude *A*, which is in accordance with the weak principle *W* but not with the strong principle *Z*, which recommends $\neg A$. But principle *Z* is actually the correct principle of induction and principle *W* is not, and attitude *A* is impermissible.

In this kind of case, the pluralist about epistemic permissibility can say that although Eric is *ideally* obligated to believe that principle *Z* is correct and that attitude *A* is impermissible (e.g. because principles of rationality like *Z* are knowable a priori), there is nonetheless a sense in which it is permissible or even obligatory for Eric to believe – on the basis of his research – that principle *W* is correct and that attitude *A* is obligatory. Eric's belief that attitude *A* is obligatory is *non-ideally* permissible, even though it is *ideally* impermissible. Can adopting this kind of pluralist account of epistemic permissibility help the fallibilist to avoid the No Way Out argument?

In order to see how adopting the view there is more than one kind of epistemic permissibility could help the fallibilist to avoid the No Way Out argument, we need to look at what sort of anti-akrasia principles that those who adopt this kind of dual-level view of epistemic permissibility will endorse. The anti-akrasia principle outlined in the previous section operates across a single kind of epistemic permissibility. If the fallibilist holds that there are two different kinds of epistemic permissibility, then there are four variants of the original Anti-Akrasia that she can commit to. Let us use $IP(A)$ to indicate that an attitude *A* is *ideally permissible*, $NP(A)$ to indicate an attitude *A* is *non-ideally permissible*, $IO(A)$ to indicate that an attitude *A* is *ideally obligatory*, and $NO(A)$ to indicate an attitude *A* is *non-ideally obligatory*. We can formulate the possible variants of Anti-Akrasia as follows:

> **(AA1) Uniform Ideal Anti-Akrasia**: $\neg\exists S\,\exists A(IP_S(A \wedge B(\neg IP_S(A))))$
> (There is no situation S in which an agent is ideally permitted to jointly believe that attitude *A* is ideally impermissible in S and to adopt attitude *A* in S)
>
> **(AA2) Non-Uniform Ideal Anti-Akrasia**: $\neg\exists S\,\exists A(IP_S(A \wedge B(\neg NP_S(A))))$
> (There is no situation S in which an agent is ideally permitted to jointly believe that attitude *A* is non-ideally impermissible in S and to adopt attitude *A* in S)
>
> **(AA3) Non-Uniform Non-Ideal Anti-Akrasia**: $\neg\exists S\,\exists A(NP_S(A \wedge B(\neg IP_S(A))))$
> (There is no situation S in which an agent is non-ideally permitted to jointly believe that attitude *A* is ideally impermissible in S and to adopt attitude *A* in S)
>
> **(AA4) Uniform Non-Ideal Anti-Akrasia**: $\neg\exists S\,\exists A(NP_S(A \wedge B(\neg NP_S(A))))$
> (There is no situation S in which an agent is non-ideally permitted to jointly believe that attitude *A* is non-ideally impermissible in S and to adopt attitude *A* in S)

Given the distinction between ideal and non-ideal permissibility, there are eight possible variants of Strong Normative Fallibilism that the fallibilist with a dual-level view of epistemic permissibility could adopt. In order to see how the above anti-akrasia principles relate to Strong Normative Fallibilism, we need to distinguish between these variants of Strong Normative Fallibilism, just as we did for Anti-Akrasia. We can formulate these variants as follows:

(SNF1) $\exists S \exists A(IO_S(A) \wedge IP_S(B(\neg IP_S(A))))$

(SNF2) $\exists S \exists A(IO_S(A) \wedge NP_S(B(\neg IP_S(A))))$

(SNF3) $\exists S \exists A(NO_S(A) \wedge IP_S(B(\neg IP_S(A))))$

(SNF4) $\exists S \exists A(NO_S(A) \wedge NP_S(B(\neg IP_S(A))))$

(SNF5) $\exists S \exists A(IO_S(A) \wedge IP_S(B(\neg NP_S(A))))$

(SNF6) $\exists S \exists A(IO_S(A) \wedge NP_S(B(\neg NP_S(A))))$

(SNF7) $\exists S \exists A(NO_S(A) \wedge IP_S(B(\neg NP_S(A))))$

(SNF8) $\exists S \exists A(NO_S(A) \wedge NP_S(B(\neg NP_S(A))))$

It's not necessary for us to work through each of these variants of Strong Normative Fallibilism. We just need to know which of them are ruled out by the anti-akrasia principles AA1 - AA4 if we assume DIR. Non-ideal permissibility does not entail ideal permissibility and vice versa, and so each of the four anti-akrasia principles only entails the negation of one variant of Strong Normative Fallibilism. These are as follows:

$$(AA1 \wedge DIR) \rightarrow \neg SNF1 \qquad (AA2 \wedge DIR) \rightarrow \neg SNF5$$

$$(AA3 \wedge DIR) \rightarrow \neg SNF4 \qquad (AA4 \wedge DIR) \rightarrow \neg SNF8$$

The Anti-Akrasia principle that Titelbaum appeals to (AA) seems to correspond with AA1 above. AA1 and DIR entail that SNF1 is false: they entail that it cannot be the case that attitude *A* is ideally obligatory, while it is ideally permissible for the agent to believe that *A* is ideally impermissible. In the case above, Eric is ideally obligated to believe that attitude *A* is impermissible. And so, by AA1, it is ideally impermissible for Eric to adopt attitude *A*.

It is likely that fallibilists who endorse a distinction between ideal and non-ideal permissibility will accept the strong normative fallibilist principle SNF2 and reject the strong normative fallibilist principle SNF1. They will claim that agents like Eric cannot have *ideally permissible* false beliefs what attitudes are *ideally permissible*, but they can nonetheless have *non-ideally permissible* false beliefs about what attitudes are *ideally permissible*. In other words, Eric is *non-ideally* permitted to believe – on the basis of his research – that principle *W* is correct and that he *ideally* ought to have attitude *A*, even though he *ideally* ought to believe that principle *Z* is correct and that attitude *A* is impermissible. If the fallibilist adopts SNF2 instead of SNF1 then she can accept the anti-akrasia principle AA1. Because she rejects SNF1 we cannot generate a contradiction with AA1, since the fallibilist who accepts SNF2 can accept that it is never ideally permissible to believe that an attitude is ideally impermissible when the attitude is ideally obligatory.

If the fallibilist accepts SNF2 and rejects SNF1, then the more relevant anti-akrasia principle to consider is AA3. SNF2 says that agents can be *non-ideally* permitted to believe that *A* is *ideally* impermissible even though *A* is *ideally* permissible. And principle AA3 says that an agent is never *non-ideally* permitted to believe that *A* is *ideally* impermissible and to adopt attitude *A*.

If AA3 entailed ¬SNF2, then we could formulate a version of the No Way Out argument for those who accept SNF2 and reject SNF1. But AA3 does not entail ¬SNF2. It entails ¬SNF4.

Someone who denies SNF4 must say that it cannot jointly be the case that attitude *A* is *non-ideally obligatory* while it is *non-ideally permissible* for the agent to believe that *A* is *ideally impermissible*. For example, suppose that having attitude *A* is non-ideally obligatory for Eric in the case above. Then, by denying SNF4, we must say that it is not *non-ideally permissible* for Eric to believe that *A* is ideally impermissible. Since Eric is *non-ideally permitted* to believe that *A* is *ideally obligatory* in his case, the case is consistent with a denial of SNF4. Indeed: SNF4 seems like a principle that the fallibilist should reject. It seems plausible that an attitude cannot be non-ideally obligatory if it is non-ideally permissible for the agent to believe that the attitude in question is ideally impermissible. Unlike ¬SNF2, ¬SNF4 is compatible with the fallibilist's claim that agents can make non-ideally permissible mistakes about ideal permissibility.

So the dual-level fallibilist can claim that her position involves a commitment to SNF2 but not to SNF4 or SNF1. In doing so, she can accept anti-akrasia principles AA1 and AA3 without this generating a dual-level version of the No Way Out argument.

Given this, it might seem that distinguishing between ideal and non-ideal permissibility can help the fallibilist to overcome the original No Way Out argument (if we assume that AA corresponds to AA1). But a problem remains for the fallibilist. The original No Way Out argument does not undermine fallibilists who endorse non-uniform anti-akrasia principles like AA3. But a version of the No Way Out argument can be formulated for all *uniform* anti-akrasia principles like AA1 and AA4 and *uniform* strong normative fallibilist theses like SNF1 and SNF8. We can just reject the form of fallibilism – SNF1 – that conflicts with AA1. But it seems like fallibilists who reject SNF1 should be committed to SNF8: the claim that an agent can be *non-ideally obligated* to adopt attitude *A*, but *non-ideally obligated* to believe that attitude *A* is *non-ideally impermissible*. And yet this conflicts with the plausible anti-akrasia principle AA4. AA4 says that it's *non-ideally impermissible* to adopt attitude *A* while believing that *A* is *non-ideally impermissible*.

Why should the fallibilist be committed to SNF8? It seems that, just as we can be mistaken about what we are ideally permitted to believe, we can also be mistaken about what we are permitted to believe given our cognitive limitations. But AA4 and the Deontic Inference Rule jointly entail ¬SNF8. If we deny SNF8 then we must deny that an attitude is non-ideally obligatory while, at the same time, the agent is non-ideally permitted to believe that the attitude is non-ideally impermissible. But it seems like we could be mistaken about what attitudes are non-ideally obligatory. And so if fallibilists adopt a distinction between ideal and non-ideal permissibility, the No Way Out argument simply re-arises at the level of non-ideal permissibility.

In order to avoid the non-ideal variant of the No Way Out argument, the fallibilist could argue that there are infinitely many kinds of epistemic permissibility and infinitely many kinds of non-uniform anti-akrasia principles. If the fallibilist does this then they can deny SNF8 just as they denied SNF1, but argue that there is another *even less ideal* form of epistemic permissibility than non-ideal permissibility: let's call it super non-ideal permissibility. And even though we cannot be non-ideally permissibly mistaken about non-ideal permissibility, we can be super non-ideally permissibly mistaken about non-ideal permissibility. But it seems desirable to avoid positing infinitely many kinds of epistemic permissibility in order to avoid the No Way Out argument.

Alternatively, the fallibilist could argue that there is some level at which epistemic permissibility 'bottoms out': a level at which an agent cannot *in any sense* have permissible false beliefs about what she is permitted to believe at that very level. At this 'base level' of epistemic permissibility, an agent could not be permissibly mistaken about what her obligations are. If the fallibilist were to accept this kind of view, she would need to explain why an agent cannot be permissibly mistaken about what her obligations are at some level of epistemic permissibility, and how this can be consistent with fallibilism. In the next two sections I will set aside dual-level views, but I will attempt to explain why the view that an agent cannot be permissibly mistaken about what her obligations can be consistent with thoroughgoing fallibilism. This may offer some support to those with a dual-level view who wish to argue that there is some level at which epistemic permissibility 'bottoms out'. But since this solution can be appealed to even if there is only one kind of epistemic permissibility, it will also remove the incentive to adopt a multi-level view of permissibility.

3 Bottom Up, Mismatch, and Top Down Views

Appealing to different kinds of epistemic permissibility seems to get around the No Way Out problem. But it does so ate the cost of either positing a regress of epistemic permissibility, or positing that there must be some level or kind of epistemic permissibility about which we must be normatively infallible. In this section, I will explore a possible solution to the No Way Out problem for fallibilists who want to commit to there being a single kind of epistemic permissibility.

To begin with, let us consider a slight variant of a case given in Titelbaum (p. 277-8, 2015).[11] Suppose that Jane initially believed that a given proposition p is not both true and false, and that it would be impermissible for her to believe that it's both true and false. In other words, she believed both $\neg(p \wedge \neg p)$ and $\neg P(B(p \wedge \neg p))$. But Jane has been learning logic under a brilliant but controversial logician, who has convinced her that p is in fact both true and false at the same time. In fact, his arguments are so convincing that Jane comes to believe both $(p \wedge \neg p)$ and that that it is *impermissible* for her to believe $\neg(p \wedge \neg p)$ in her current situation.

Titelbaum asks us to consider two questions. First, is there any way that we can fill in the background facts about Jane's situation so that Jane's belief that it's impermissible to believe $\neg(p \wedge \neg p)$ is itself a an *obligatory* belief? Second, is there any way to fill in the facts about Jane's situation so that Jane is *permitted* to believe $(p \wedge \neg p)$ in this kind of case?[12]

Titelbaum uses the different combinations of answers to these questions to distinguish between three positions that one can have on the relationship between first-order attitudes and attitudes about epistemic permissibility. These are the *Bottom Up* view, the *Mismatch* view, and the *Top Down* view. The Bottom Up view answers 'no' to both of these questions: Jane cannot be obligated

[11] The case that Titelbaum considers involves an agent who adopts logically inconsistent beliefs that she believes are *permissible*. But only cases in which an agent believes that an attitude is *impermissible* or *obligatory* that create problems for fallibilism. And so I adjust the case to one that involves believing that an attitude is impermissible.

[12] Titelbaum asks whether Jane's belief about epistemic permissibility can be *permissible* rather than obligatory. However, since the problematic cases will involve *obligatory* rather than merely permissible beliefs about epistemic permissibility, I will formulate the three views in relation to *obligatory* beliefs about epistemic permissibility.

to have a mistaken belief about what is epistemically impermissible in her current situation, and she cannot permissibly believe $(p \wedge \neg p)$ in her current situation. The Mismatch view answers 'yes' to the first question and 'no' to the second question: Jane can be obligated to believe that believing $\neg(p \wedge \neg p)$ is impermissible on the basis of the arguments and testimony of her logic teacher, but this does not make it rationally permissible for her to believe $(p \wedge \neg p)$.[13] The Top Down view answers 'yes' to the first question and 'yes' to the second question: Jane can be obligated to believe that believing $\neg(p \wedge \neg p)$ is epistemically impermissible, and she can permissibly believe $(p \wedge \neg p)$ in cases like this. So the Bottom Up, Mismatch, and Top Down views say the following:

	$O_S(B(\neg P_S(B(\neg(p \wedge \neg p)))))$	$P_S(B(p \wedge \neg p))$
Bottom Up	✗	✗
Mismatch	✓	✗
Top Down	✓	✓

It should be noted that the Bottom Up, Mismatch, and Top Down views can all be formulated as either *general* principles, or as *case-specific* principles. The general Bottom Up view says that there is *no* case in which an agent can have a permissible mistaken believe that some attitude is impermissible when it is not. The case-specific Bottom Up view just says that there are *some* cases in which an agent cannot have a permissible mistaken belief that some attitude is impermissible when it is not. Fallibilists do not need to commit to a general version of any of these principles, since fallibilists only claim that there are *some* cases in which an agent can be obligated to adopt a false normative belief: they do not need to claim that this is true in all cases like Jane's. And so they only need to be committed to either the Top Down or Mismatch views in *some* cases. Given this, I will be discuss *case-specific* versions of these three principles here.

Those who adopt the Bottom Up view in Jane's case can accept that Anti-Akrasia is a principle of epistemic permissibility because, when it comes to cases like Jane's, they will reject the key commitment of Strong Normative Fallibilism: the joint possibility of Jane being obligated to adopt attitude A and her being obligated to believe that it's impermissible for her to adopt attitude A. If it is impermissible for Jane to adopt attitude $(p \wedge \neg p)$ in her current situation S then, according to the Bottom Up view, it is impermissible for her to believe that believing $(p \wedge \neg p)$ is permissible in S. This is called the 'Bottom Up' view because it says that if it's not permissible for you to adopt the 'first-order' attitude A, this means that it's not permissible for you to adopt the 'higher-order' attitude $B(\neg P(A))$. In other words, what is permissible at the lower level affects what is permissible at the higher level. Fallibilists cannot think that the Bottom Up view is true in *all* cases like Jane's.

Those who adopt the Mismatch view in Jane's case can accept Strong Normative Fallibilism because they will reject either the Anti-Akrasia principle or they will reject the Deontic Inference Rule in Jane's case. According to the Mismatch view, Jane can be obligated to believe – on the basis of her professor's arguments and testimony – that it is impermissible for her to believe $\neg(p \wedge \neg p)$ and, at the same time, to believe she is not permitted to believe $(p \wedge \neg p)$. This conflicts

[13] We can imagine a less plausible Mismatch view that answers no to the first question and yes to the second question. But I won't consider this kind of Mismatch view here.

with Anti-Akrasia because Jane is obligated to believe that it is impermissible for her to fail to believe $(p \wedge \neg p)$ and yet she is also obligated to fail to believe $(p \wedge \neg p)$. So those who adopt the Mismatch view believe that the Deontic Inference Rule is true, then they must believe that Anti-Akrasia is not a genuine epistemic principle in Jane's case.

Finally, those who adopt the Top Down view in Jane's case can accept Anti-Akrasia in this case. This is because, like those who adopt the Bottom Up view, those who accept the Top Down view reject the Strong Normative Fallibilism in Jane's case. The Top Down view says that Jane can be *obligated* to believe – on the basis of the arguments and testimony of her logic professor – that it is impermissible for her to believe $\neg(p \wedge \neg p)$. But it also says that she is permitted to believe $(p \wedge \neg p)$ in these circumstances. The Top Down view therefore retains Anti-Akrasia. But in doing so it must reject Strong Normative Fallibilism.

In the case described above, the Top Down view says that it is permissible for Jane to believe $(p \wedge \neg p)$, even if we suppose that she *would have been* impermissible for her to believe $(p \wedge \neg p)$ prior to hearing her professor's arguments and testimony. This is called the 'Top Down' view because it says that if it's obligatory or permissible for you to adopt the 'higher-order' attitude $B(\neg P(A))$, this can mean that it's permissible for you to adopt the 'first-order' attitude $\neg A$ in S. In other words, what is permissible at the higher level affects what is permissible at the lower level.

Many fallibilists will want to maintain that, contrary to what the Bottom Up view says, agents can be permissibly mistaken about whether an attitude is obligated, permitted, or forbidden even in a case like Jane's. But they will presumably want to avoid saying, as the Mistmatch view does, that either Anti-Akrasia or the Deontic Inference Rule must be false in these cases. I believe that, although it rejects Strong Normative Fallibilism, the Top Down view can offer the fallibilist a way of saying that we can be permissibly mistaken about which attitudes we are obligated to believe, without saying that Anti-Akrasia or the Deontic Inference Rule is false. In the next section I will outline what I take to be the most plausible motivation for the Top Down view, and will place it within a more general account of what the competing considerations are in cases like Jane's. I will then respond to the objection that, because the Top Down view is incompatible with Strong Normative Fallibilism, the view is ultimately incompatible with thoroughgoing fallibilism.

4 Level-Connection Principles and the Top Down View

In order to outline what I take to be the most plausible motivation for the Top Down view, I think that it would be useful to reformulate some of the discussion that has occurred so far in terms of *epistemic reasons*. This will let us uncover some of the principles motivating the Bottom Up, Mismatch, and Top Down views. The principles about epistemic reasons that are relevant to cases like Jane's are so-called 'level-connection' principles: principles that connect an agent's reasons for having an ordinary first-order attitude A with her reasons for having higher-order attitudes: that is, her reasons for having attitudes about that first-order attitude.

It seems that Jane – like everyone else – has epistemic reasons to avoid believing in logical falsehoods. But it also seems that Jane has epistemic reasons to believe that she is *obligated* to believe in logical falsehoods in her particular case. Do Jane's reasons to avoid believing logical falsehoods give her any reasons to believe that it's *impermissible* for her to believe logical falsehoods? Do her reasons to believe that she is *obligated* to believe logical falsehoods in give her reasons to believe in those logical falsehoods? If we answer 'yes' to either question then this suggests that there is some sort of level-connection principle between an agent's reasons to adopt an attitude, and her reasons to adopt an attitude about the permissibility of that attitude.

In what follows I will assume that epistemic reasons are pro tanto considerations in favor of adopting an attitude that may come in different strengths. I will remain neutral on what constitutes an epistemic reason: it may be an agent's evidence, her other attitudes, facts about the world, and so on. I will, however, assume a close link between reasons and rational obligations: namely, that an attitude is obligatory if and only if we have sufficient reasons to adopt that attitude, and that an attitude is impermissible if and only if we have sufficient reasons against adopting that attitude. What counts as 'sufficient' reasons may be relative to the options available to the agent: for example, we may think that an attitude is obligatory if and only if it is the attitude that you have the most reasons to adopt. But if this view is correct then we could – in principle – have no reason to adopt a given attitude and yet be obligated to adopt that attitude (because we have reasons *against* adopting all of the other attitudes available to us). As a result, epistemic reasons can alter the deontic status of a given attitude: gaining epistemic reasons for adopting an attitude can take that attitude from being impermissible to permissible or even obligatory, and losing epistemic reasons for adopting an attitude can take that attitude from being obligatory to permissible or even impermissible. I am therefore going to be assuming that if we are obligated to satisfy principles like Anti-Akrasia in a given case, it's because we have epistemic *reasons* to avoid having attitudes that are akratic. In other words, I assume that the following principle is true:

> **Reasons Anti-Akrasia (RAA):** An agent has epistemic reasons against jointly believing that attitude A is impermissible in S and adopting attitude A in S

Let us say that an agent has *conclusive* reasons to adopt attitude A if and only if her reasons in favor of adopting attitude A are sufficiently strong for her to be epistemically obligated to adopt attitude A. Then we can transform Reasons Anti-Akrasia into the original Anti-Akrasia principle by claiming that the agent's reasons against jointly believing that attitude A is impermissible and adopting attitude A are always conclusive: she always has more reasons to avoid being epistemically akratic than she does to jointly adopt some attitude A and the belief that A is impermissible.

There is little consensus about how we should formulate and justify level-connection principles in epistemology. But one way we can argue for the claim that an agent's reasons for adopting some first-order attitude A give her reasons against believing that A is permissible – and that her reasons for believeing that A is impermissible give her reasons adopting attitude A – is that she has independent reasons to avoid having an overall belief state that is akratic, as RAA states.

To see how Reasons Anti-Akrasia supports level-connection principles, suppose that p and q are inconsistent with one another and that agents are obligated to avoid adopting inconsistent beliefs. Then if an agent had a reason for believing p she would also have a reason for disbelieving q, insofar as she has reasons to avoid believing both p and q at the same time. Similarly, if you have reasons to avoid being epistemically akratic, then your having a reason to adopt attitude A will mean that you also have a reason to believe that attitude A is permissible. And your having a reason to believe that adopting attitude A is impermissible will mean that you also have a reason to not adopt attitude A. In other words, Reasons Anti-Akrasia behaves like a coherence norm between higher-order attitudes and first-order attitudes.[14]

If this is correct then Reasons Anti-Akrasia can be used to justify level-connection principles. If we have reasons to avoid jointly adopting A and believing that A is impermissible, then a reason *for* adopting A will provide us with a reason *against* believing that A is impermissible, and a reason *for* believing that A is impermissible with provide us with a reason *against* adopting A. These are reasons-based level connection principles between first-order and higher-order attitudes.

Although I don't claim that the argument from Reasons Anti-Akrasia given above is the sole or even the primary justification of level-connection principles, it is the one that I will adopt here, since Titelbaum accepts that we ought to avoid having an overall epistemic state that is akratic, and this view will commit us to level-connection principles existing between our ordinary first-order attitudes and our higher-order attitudes about which first-order attitudes are permissible or not in a given situation. I won't attempt to offer further justifications of level-connection principles here.

We can now formulate the two different level-connection principles that can exist between the reasons that an agent has for adopting first-order attitudes and for adopting higher-order attitudes (attitudes about the permissibility of those first-order attitudes) as follows:

> **Reasons Down (RD):** If S is a situation in which (i) an agent has some epistemic reasons for believing that it is impermissible for her to adopt attitude A in S, and (ii) the agent comes to believe, on the basis of these reasons, that it is impermissible for her to adopt attitude A in S, then the agent has some epistemic reasons against adopting attitude A in S
>
> **Reasons Up (RU):** If S is a situation in which (i) an agent has some epistemic reasons for adopting attitude A in S, and (ii) the agent, on the basis of these reasons, comes to adopt attitude A in S, then the agent has some epistemic reasons against believing that A is impermissible in S

If we think that Reasons Anti-Akrasia is true, and that an agent can have *independent* reasons to adopt a given first-order attitude A (that is: reasons not generated by her reasons for believing

[14] Notice that having reasons to adopt attitude A doesn't mean that you will have reasons of the *same strength* for believing that attitude A is permissible, and having reasons to believe that that adopting attitude A is impermissible does not mean that you will have reasons of the *same strength* for not adopting attitude A. All that matters is that your having reasons for believing the first claim mean that you have *some* reasons for believing the second, even if they are not of the same strength. It may also be the case that your reasons for adopting attitude A only give you reasons to believe that adopting attitude A is permissible if, on the basis of your reasons for adopting attitude A, you do in fact come to adopt attitude A. Nothing that I say here is inconsistent with this sort of view.

that she is obligated or permitted to adopt attitude *A*), then it seems we should endorse a level-connection principle like Reasons Up. And if we think that Anti-Akrasia is true, and that an agent can have *independent* reasons to have the higher-order belief that a first-order attitude *A* is impermissible (that is: reasons not generated by her reasons against adopting attitude *A*), then it seems we should endorse a level-connection principle like Reasons Down. And if we think that it is possible for an agent to have both independent reasons to adopt a first-order attitude, and independent reasons to have a higher-order belief that a first-order attitude is impermissible, then it seems we should endorse both the 'up' and 'down' level-connection principles formulated above.

By accepting both principles, we can avoid committing to one particular account of what should happen in cases like Jane's. If the relevant case is one in which an agent has more overall reasons to adopt attitude *A* and to avoid being epistemically akratic than she has overall reasons to believe that attitude *A* is impermissible, then we can conclude that she is obligated to believe that attitude *A* is impermissible and that she is not permitted to adopt attitude *A*. In other words, we can adopt a Bottom Up view in that case. If, on the other hand, the relevant case is one in which the agent has more overall reasons to adopt attitude *A* and to believe that attitude *A* is impermissible than she has overall reasons to avoid being epistemically akratic, then we can conclude that she is obligated to be epistemically akratic. In other words, we can adopt a Mismatch view in that case. And if the relevant case is one in which the agent has more reasons to believe that *A* is impermissible and to avoid being epistemically akratic than she has reasons to adopt attitude *A*, then we can claim that she is permitted to adopt attitude *A* and she is not obligated to believe that attitude *A* is impermissible. In other words, we can adopt a Top Down view in that case.

We can use Reasons Down to motivate adopting a Top Down view in cases like Jane's. If an agent has *conclusive* reasons to believe that *A* is impermissible then, because she also has *conclusive* reasons to avoid being epistemically akratic (if we assume that the original Anti-Akrasia principle is correct), she cannot have *conclusive* reasons to adopt attitude *A*. Her reasons for believing that *A* is impermissible and for avoiding being epistemically akratic give her reasons against adopting attitude *A*.All that the fallibilist who accepts the original Anti-Akrasia principle requires is that there are some cases in which an agent can have conclusive reasons to believe that an attitude that attitude *A* is impermissible, where *A* is an attitude that she *would* have had conclusive reasons to adopt were it not for her conclusive reasons to believe that it is impermissible.[15] If this is the case, then she will say that although the agent *would* have had conclusive reasons to adopt attitude *A*, her conclusive reasons to believe that *A* is impermissible mean that she cannot have conclusive reasons to adopt attitude *A*.[16] Those who adopt this view can accept Anti-Akrasia without having to say that an agent can never be mistaken about what she has conclusive reasons to believe, and so it offers the fallibilist a plausible way out of the No Way Out argument.

[15] This need not happen in all cases: the fallibilist can accept that there are some cases in which an agent has stronger reasons to adopt *A* and avoid akrasia (Bottom Up cases), or weaker reasons to avoid akrasia (Mismatch cases).

[16] Although I have formulated the principles here in terms of epistemic reasons, we can make the same point with epistemic obligations. We can say that if an agent is obligated to believe that *A* is impermissible then, because she is also obligated to avoid being epistemically akratic, her obligatory belief that *A* is impermissible must mean that although it may true that attitude *A* *would* have been obligatory, it is now merely permissible, or even impermissible.

To look at how this works in the example given earlier, suppose we think that before Jane heard anything from her professor, she had conclusive reasons for believing $\neg(p \wedge \neg p)$. After she hears the arguments of her professor Jane has conclusive reasons for believing – on the basis of her professor's testimony – that she has conclusive reasons *against* believing $\neg(p \wedge \neg p)$. If we accept the level-connection principle Reasons Down, this means that after she hears from her professor, Jane must have some reasons for believing $(p \wedge \neg p)$ because she has reasons to avoid being akratic and she has reasons believe $\neg P_S(B(\neg(p \wedge \neg p)))$. And since Jane's reasons to avoid being akratic are conclusive and her reasons to believe $\neg P_S(B(\neg(p \wedge \neg p)))$ are also conclusive, Jane can no longer have conclusive reasons to believe $\neg(p \wedge \neg p)$.

The Top Down view says that Jane is obligated to believe $\neg P_S(B(\neg(p \wedge \neg p)))$. Let us suppose that she is also obligated to avoid being akratic. This means that even if Jane *was* obligated to believe $\neg(p \wedge \neg p)$ prior to her professor's testimony, she is not obligated to believe $\neg(p \wedge \neg p)$ after hearing his testimony. This is consistent with the Top Down view, since the Top Down view also says that Jane is permitted to believe $(p \wedge \neg p)$ after receiving the professor's testimony.[17] So, if the Top Down view is correct in this case, Jane can be obligated to believe things about epistemic impermissibility that are *in some sense* mistaken – since Jane's belief that $\neg(p \wedge \neg p)$ was obligatory prior to her receiving testimony that it is impermissible – without thereby violating the Anti-Akrasia principle. So the Top Down view – motivated by the level-connection principle Reasons Down – seems to offer the fallibilist a viable way out of the No Way Out argument.

The most pressing objection to this response, however, is that the Top Down view cannot support a genuinely fallibilist view about epistemic normativity - i.e. a view which says that agents can be permissibly *mistaken* about whether an attitude is permissible or not in a given case. Like the Bottom Up view, the Top Down view rejects Strong Normative Fallibilism. Strong Normative Fallibilism says that there exists a situation S in which an attitude A is obligatory, but it is permissible for the agent in S to believe that A is impermissible in S. But, according to the Top Down view, if an agent is obligated to believe that attitude A is impermissible in her current circumstances, and if she is also obligated to avoid epistemic akrasia, then this means that attitude A is not obigatory in her current circumstances. In other words, her higher-order beliefs about epistemic permissibility can affect the normative status of her first-order beliefs. So in what sense can an agent's belief that A is impermissible in her current circumstances be a *mistake*?

I believe that fallibilists can reject Strong Normative Fallibilism while still retaining a plausible degree of fallibilism about epistemic permissibility. There are two key respects in which an agent's belief that an attitude A is impermissible can be *mistaken* even if – as the Top Down view states – her having an obligatory belief that A is impermissible means that A is no longer obligatory.

First, an agent can still have false or mistaken beliefs about the normative status of A even if she has a permissible belief that A is impermissible. For example, she may falsely believe that A is

[17] If Jane's reasons for failing to be akratic are conclusive and her reasons for believing $\neg(p \wedge \neg p)$ are conclusive, then she cannot have conclusive reasons for believing $(p \wedge \neg p)$. This is the Bottom Up analysis of Jane's case. And if Jane's reasons for believing $\neg(p \wedge \neg p)$ are conclusive and her reasons for believing $\neg P_S(B(\neg(p \wedge \neg p)))$ are conclusive, then she cannot have conclusive reasons for avoid akrasia. This is the Mismatch analysis of Jane's case.

impermissible when *A* is in fact *permissible but not obligatory*. According to the Top Down view, her having conclusive reasons to believe that *A* is impermissible does mean that she cannot have conclusive reasons to adopt attitude *A*. But this need not mean that Jane has conclusive reasons *against* adopting *A*. And so her belief that *A* is impermissible (i.e. that she has conclusive reasons against adopting *A*) would still be mistaken in this case. In other words, her having a conclusive reasons to believe that *A* is impermissible merely reduces the severity of her mistake in this sort of case, but it does not prevent her from making one.

To see another important respect in which the agent's belief can be mistaken, we need to consider what it is that the agent believes when she believes that *A* is impermissible. She might believe something like 'given that I have reasons to believe that *A* is impermissible, *A* is impermissible', or she might believe something like 'attitude *A* is impermissible, and it would be impermissible regardless of what reasons I have to believe that *A* is impermissible'. When we have beliefs about normative matters, it often seems to be the latter kind of belief that we have. For example, suppose that I believe that it is impermissible for me to torture one person in order to save three, because I have good reasons to believe that torture is always wrong. My belief is not merely that it is wrong for me to torture one person *given my permissible attitudes about torture*. My belief is that it is wrong for me to torture one person because torture is always wrong regardless of my beliefs. If it turns out that torture is only wrong in my current circumstances because I have good reasons to believe that torture is always wrong, I will falsely believe that torture is wrong *independent of my permissible attitudes about torture*.[18] The Top-Down view says that if an agent is obligated to believe that *A* is impermissible then, given that this is an obligatory belief, the agent is not obligated to adopt attitude *A*. But if an agent in these circumstances were to believe '*A* is impermissible independent of my permissible attitudes about *A*' then she would believe something false.[19]

So although the Top Down view is inconsistent with Strong Normative Fallibilism, rejecting this principle in favor of the Top Down view lets us embrace a degree of fallibilism about normative matters that should be acceptable to most fallibilists. The Top Down view therefore seems to offer a plausible way out of the No Way Out argument for those fallibilists who accept Weak Normative Fallibilism and do no wish to reject the Anti-Akrasia principle.

However, the Top Down view can only be used to support fallibilism if there are indeed cases in which an agent can be obligated to believe that an otherwise obligatory attitude *A* is impermissible. Titelbaum offers an additional argument against the view that an an agent can be obligated to believe that an otherwise obligatory attitude *A* is impermissible. We must consider this additional argument in order to complete this defense of fallibilism against the No Way Out argument.

[18] This mirrors, to some extent, the distinction between subjective and objective oughts, though the latter are generally used only in cases of empirical uncertainty.

[19] This position is to be consistent with the view mentioned at the end of section 2, which says that epistemic permissibility 'bottoms out' at some level. But note that the agent's obligatory belief that *A* is impermissible does not guarantee knowledge that, given her obligatory belief that *A* is impermissible, she is not obligated to adopt attitude *A*, since the agent may lack any access to the fact that her belief that *A* is impermissible is itself an obligatory belief.

5 The Self-Undermining Argument

Normative fallibilists are committed to there being cases in which an agent has permissible but mistaken beliefs about which attitudes are permissible or impermissible in her current situation (Weak Normative Fallibilism). We can appeal to the Top Down view to avoid the No Way Out argument for strong fallibilism. But this only works if there are situations in which an agent has independent reasons to believe that it is impermissible for her to believe A, even though – if it were not for these independent reasons – she would have been obligated to adopt attitude A.

Titelbaum (p. 270, 2015) presents a challenge to this view that there can be obligatory false beliefs of the form: 'A is impermissible in S'. As we will see, for the argument of this section to succeed, we need there to be cases in which an agent is obligated to believe that A is impermissible even though, were it not for this belief, A would be obligatory. If we can be obligated to have beliefs of this form then, Titelbaum claims, it must be because there are epistemic principles which say that false beliefs of the form 'A is impermissible in S' can be obligatory. Titelbaum offers an example of such a principle, 'Testimony', which is as follows:[20]

> **Testimony (T)** If an agent's situation includes testimony that x, the agent is rationally permitted and required [obligated] to believe that x

For a principle like Testimony to be true it's going to have to be hedged in various respects: Testimony certainly seems like it's not going to be true for all propositions.[21] But if fallibilists are committed to the claim that we can be obligated to believe that an attitude that would otherwise be obligatory is impermissible, then there must be *some* principle or set of principles that say we can be obligated to have beliefs of this sort. And Testimony is an example of an epistemic principle that can result in our being obligated to have this kind of belief about epistemic obligations. So let us assume that something like Testimony applies to at least certain propositions in certain situations. Let us assume that t is the kind of proposition that the Testimony principle applies to:

> t: if an agent's situation includes testimony that x, the agent is rationally forbidden to believe that x

Titelbaum (ibid.) argues that if fallibilists accept Anti-Akrasia and a testimony principle like Testimony, which that says agents can be obligated to believe t, then they are faced with the following problem if an agent receives testimony that t:

> 'By Testimony, the agent in this situation is permitted an overall state in which she believes t. So suppose the agent is in that rationally permitted state. Since the agent believes t, she believes that it's rationally impermissible to believe testimony. She learned t from testimony, so she believes that belief in t is rationally forbidden in

[20] I have copied Titelbaum's formulation of Testimony here. But we can replace 'permissible and required [obligated]' with just 'required [obligated]' in Testimony, since mere permissibility is not sufficiently strong for Titelbaum's argument to go through, and we are assuming that obligation entails permissibility anyway.

[21] For example, my evidence could include testimony that p and testimony that $\neg p$. Then, according to Testimony, I would be obligated to believe both p and not p, and surely this cannot be correct.

> her situation. But now her overall state includes both a belief in *t* and a belief that believing *t* is rationally forbidden. By the Akratic Principle, the agent's state is rationally impermissible, and we have a contradiction. The Akratic Principle entails that Testimony is not a true rule of rationality.'

The worry is that in order to satisfy Testimony, the agent is obligated to believe that *t*. But if the agent believes *t* then she will believe that it is impermissible for her to believe *t*, since her situation includes testimony that *t*. So if the agent retains her belief in *t* then her attitude state will contain *t* and the belief that she is not permitted to believe *t*, which violates the Anti-Akrasia principle. If, on the other hand, the agent believes Testimony and not *t* then she also violates the Anti-Akrasia principle, because if she believes Testimony then she believes she is not permitted to fail to believe *t*. So the only non-akratic belief state in this case is one in which the agent believes neither Testimony nor *t*, and that belief state is not permitted by Testimony. This means that the agent cannot satisfy Anti-Akrasia without violating Testimony, and vice versa:

	Anti-Akrasia	Testimony
B($t \wedge T$)	✗	✓
B($t \wedge \neg T$)	✗	✓
B($\neg t \wedge T$)	✗	✗
B($\neg t \wedge \neg T$)	✓	✗

And so it seems that fallibilists need to be committed to epistemic principles that can sometimes say that we are obligated to believe that some otherwise true epistemic principle is false. But this leads to a problematic kind of self-undermining.

We can strengthen this Self-Undermining argument in three respects. First, as Titelbaum points out, we might want to restrict principles like Testimony in a way that would prevent them from undermining themselves. For example, instead of accepting Testimony, we might accept Titelbaum's restricted version of Testimony:

> **Restricted Testimony:** If an agent's situation includes testimony that *x*, the agent is rationally permitted and required [obligated] to believe that *x* – unless *x* contradicts this rule.

As Titelbaum notes, Restricted Testimony is able to avoid the Self-Undermining objection. But in doing so it prevents Testimony from being a source of an agent's obligation to believe that the Testimony principle is false. So it seems that fallibilists will need to accept some principle that is not restricted in the way that Restricted Testimony is, if she wants to be able to say that agents can have obligatory beliefs that an attitude that would otherwise be obligatory is in fact impermissible.

Second, we might think that Testimony is false because we think that there are not multiple principles of epistemic rationality: instead we think that there is an 'über rule', which maps all acts in situations onto all things considered deontic properties (i.e. for all S and for all A, it says 'in situation S the agent attitude A is obligatory/permissible/impermissible, all things considered').

And perhaps the fallibilist can avoid the Self-Undermining argument if she is committed to the right kind of epistemic über rule. Suppose, for example, that the fallibilist's über rule is as follows:

> **Über Rule (U)**: You are permitted to adopt attitude A in situation S iff there is no attitude A' in S that you have strictly more epistemic reasons to adopt than A. You are obligated to adopt attitude A in situation S iff A is the only permissible attitude in S.

This Über Rule is consistent with the Anti-Akrasia principle if it is never the case that the epistemic attitude we have most reasons to adopt is an akratic attitude.[22] Can the Über Rule be self-undermining? Let us define the 'deontic inverse' of a normative principle: if a normative principle like U says A is obligatory or permissible then its deontic inverse says A is impermissible, and if U says A is impermissible then its deontic inverse says A is obligatory. Suppose that an agent has the most epistemic reasons to believe the deontic inverse of the Über Rule (U):

> **Inverse Über Rule ($\bar{U}$)**: You are permitted to adopt attitude A in situation S iff there is some attitude A' in S that you have strictly more epistemic reasons to adopt than A. You are obligated to adopt attitude A in situation S iff A is the only permissible attitude in S

The Inverse Über Rule ($\bar{U}$) says that you ought to adopt A iff you have the fewest reasons to adopt A. Suppose that a fallibilist who is committed to U says that there is at least one situation in which an agent would have the most epistemic reasons to believe that $\bar{U}$ is the correct über rule. Therefore, in some situation S, an agent is obligated to believe $\bar{U}$ according to U. But if U says that $O_S(B(\bar{U}))$ and U says that $B(\bar{U}) \rightarrow O_S\neg(B(\bar{U}))$, then she finds herself facing an über rule variant of the original self-undermining problem. So even if fallibilists accept an über rule like U, some restriction on fallibilism does follow from the self-undermining problem: agents cannot come to believe in propositions that will undermine their own normative foundations, and this means that agents cannot be obligated to believe in the deontic inverse of that über rule. So it is not just principles like Testimony that are prey to the Self-Undermining argument: even plausible über rules like U will be prey to the Self-Undermining argument as long as they sometimes say that it is permissible for an agent to believe in the deontic inverse of themselves.

Finally, we can strengthen the Self-Undermining argument by extending the problem to cases that involve multiple principles, none of which are *self*-undermining but some or all of which are *mutually*-undermining. This is important because we might think that the fallibilist does not require self-undermining principles in order to say that we can have obligatory beliefs that an attitude is impermissible: she might accept several epistemic principles, none of which undermine *themselves* but some of which undermine *other* epistemic principles. As Titelbaum points out, we cannot appeal to these sorts of undermining principles in order to recover fallibilism. To show why, suppose that the fallibilist accepts these two principles about epistemic permissibility:

[22] The rule is also consistent with the Reasons Down-based Top Down view if, whenever we ought to believe $\neg P_S(A)$, A cannot be the attitude that we have the most reasons to adopt.

> **Restricted Testimony (RT):** If an agent's situation includes testimony that x, the agent is rationally obligated to believe that x – unless x contradicts this rule

> **Restricted Induction (RI):** If an agent's situation includes inductive evidence that x, the agent is rationally obligated to believe that x – unless x contradicts this rule

Now consider a case in which an agent gets both of the following pieces of evidence, E_1 and E_2:

> E_1 is *testimonial* evidence that P: 'if an agent's situation includes inductive evidence that x, the agent is rationally forbidden to believe that x'

> E_2 is *inductive* evidence that Q: 'if an agent's situation includes testimony that x, the agent is rationally forbidden to believe that x'

What should we say about this case? By RT, the agent is obligated to believe P and to therefore believe that RI is false. And, by RI, the agent is obligated to believe Q and to therefore believe that RT is false. So, if the fallibilist accepts both RT and RI, then the agent is obligated to believe both P and Q, and therefore to believe that RT and RI are both false. But if the agent believes P (as is obligatory by RT) then she believes it is impermissible for her to believe Q. And if she believes Q (as is obligatory by RI) then she believes that it is impermissible for her to believe P. So the agent cannot be non-akratic in this case without violating RT or RI. Although I have constructed a case that involves the two principles RT and RI, this problem generalizes to cases involving any number of principles that are mutually undermining in this way. So the Self-Undermining problem isn't restricted to cases that involve self-undermining principles: it also applies in cases where there are sets of principles that are mutually undermining.

But it is important not to overstate the problem of mutually undermining principles. Consider a case in which RI and RT are true, but where the agent only receives evidence E_1 and not evidence E_2. If this is the case then the agent could be obligated to believe P on the basis of testimony, and to therefore believe that RI is false. And her obligatory belief that RI is false will not cause her to be akratic even if RI is a true principle of rationality. The self-undermining argument will only work in cases where the propositions that we ought to believe jointly entail that we ought to disbelieve those very propositions. This is true when ought to believe in a single proposition that entails we ought not to believe that very proposition (as in Titelbaum's Testimony case and the über rule case) or when we ought to believe in a set of propositions that entail that we ought not to believe that very set of propositions (as in the RI/RT case). But it is possible to be obligated to believe that if our situation includes inductive evidence that x, then we are rationally forbidden to believe that x without believing in a set of propositions that are self-undermining or mutually undermining. This is evidenced by the case in which the agent receives E_1 and not E_2.

In order to see what responses to the self-undermining argument are available to the fallibilist, we need to give a more explicit formulation of the Self-Undermining argument. Let ϕ be a proposition that is self-undermining. In other words, ϕ trivially entails that $\neg P_S(B(\phi))$. By 'trivially entails' I mean that one ought to be such that if one believes ϕ then one believes $\neg P_S(B(\phi))$. In other words, ϕ trivially entails $\neg P_S(B(\phi))$ if and only if $O(B(\phi) \rightarrow \neg P_S(B(\phi)))$. I assume that when ϕ is equivalent to $\neg P_S(B(\phi))$ – as in Titelbaum's case – ϕ trivially entails $\neg P_S(B(\phi))$. Given this, we can formulate the self-undermining argument as follows:[23]

> (1) **Self-Undermining**: There exists a proposition ϕ such that: $O(B(\phi) \rightarrow \neg P_S(B(\phi)))$
> (2) **Normative Fallibilism**: For some ϕ there exists a situation S in which $O_S(B(\phi))$
> (3) **Anti-Akrasia**: for every situation S, $\neg P_S(A \wedge B(\neg P_S(A)))$
> (4) **No Dilemmas**: there does not exist a situation S such that $\neg P_S(A) \wedge \neg P_S(\neg A)$
>
> Contradiction from 1-4: Let S be a situation in which 1-3 are true. Suppose that the agent in S does not believe ϕ. Then she violates (2). Suppose that the agent in S believes ϕ. The agent who believes ϕ either believes $\neg P_S(B(\phi))$ or she does not believe $\neg P_S(B(\phi))$. If she believes $\neg P_S(B(\phi))$ then she violates (3). And if she does not believe $\neg P_S(B(\phi))$ then she violates (1). Therefore if 1-3 are true then $\neg P_S(B(\phi) \wedge B(\neg P_S(B(\phi))))$ and $\neg P_S \neg(B(\phi) \wedge B(\neg P_S(B(\phi))))$. And so, by (4), there does not exist such a situation S.[24]

Titelbaum would have us deny the fallibilist premise (2), and to hold instead that the true principles of rationality are such that we are never obligated to believe their deontic inverse. The fallibilist who accepts Anti-Akrasia can respond to the self-undermining argument in one of three ways: she can deny premise (4), or she can deny premise (2) and argue that this denial does not undermine the most important aspects of fallibilism, or she can deny premise (1). I will consider each of the three strategies available to her in this order. Although it is possible to reject the Anti-Akrasia principle (3) in cases of this sort, I will assume that the fallibilist is committed to principle (3) even in cases that involve self-undermining propositions like ϕ.

The first strategy that the fallibilist may wish to employ is deny premise (4) and argue that in cases of self-undermining, agents find themselves in an epistemic dilemma. If self-undermining propositions exist (as premise (1) states), we can be obligated to believe in these propositions (as premise (2) states), and we are obligated to avoid being akratic (as premise (4) states), then we must be obligated by Testimony to believe in a propositions like t and, by Anti-Akrasia, to fail to believe in that same proposition t. If there are true principles of rationality that conflict with one another as we see that Testimony and Anti-Akrasia do in this case (i.e. one says $O_S(A)$ and another says $O_S \neg(A)$, for some attitude in some state) then it is unsurprising that we could sometimes find ourselves in epistemic dilemmas. We are, after all, bound by two principles of rationality, one of which prohibits adopting some attitude in a given situation, and another which prohibits

[23] This formulation applies to self-undermining propositions but not to sets of mutually undermining propositions. The argument that captures cases that involve sets of mutually undermining propositions is slightly more complicated, but its premises are sufficiently similar to the premises of the self-undermining argument that I believe that discussion of the self-undermining argument will generalise to cases involving mutually-undermining principles.

[24] Some might want to defend an alternative to (4) that precludes dilemmas over attitudes in a single proposition, but allows for dilemmas when the attitudes in cases that involve attitudes made up of more than one belief.

failing to adopt that attitude in that situation. Such dilemmas can arise if we accept conflicting principles of rationality more generally, or if we accept rules of rationality that can sometimes produce conflicting obligations.[25] If we accept that there are true principles of rationality that produce conflicting obligations then we can retain Self-Undermining, Normative Fallibilism, and Anti-Akrasia without contradiction.

There are two important objections to this first strategy, however. The first objection is that it will commit us to the view that epistemic dilemmas are fairly widespread. Every time an agent can have an obligatory belief in a self-undermining proposition like ϕ, she will find herself in an epistemic dilemma. Even those who accept that we can *sometimes* find ourselves in epistemic dilemmas may not want to accept that epistemic dilemmas are as common as this might imply. But the second and more pressing objection is that this strategy does not recover thoroughgoing fallibilism from the Self-Undermining objection. If the fallibilist wants to be able to say that an agent can be obligated to believe in self-undermining propositions like ϕ, then surely she wants to be able to say that we are not *only* obligated to believe in self-undermining propositions in cases where we are *also* not permitted to believe in those very propositions. But this is the view we are committed to if our only response to the Self-Undermining argument is to deny premise (4). And so this response to the Self-Undermining argument is not without its difficulties.

The second strategy that the fallibilist might employ in order to avoid the Self-Undermining argument is to deny premise (2). In other words, she can deny that agents can be obligated to believe in self-undermining propositions. She will then need to show that thoroughgoing fallibilism does not need to say that we can be obligated to believe in self-undermining propositions. How much of fallibilism can we retain if we reject this premise (2)?

Notice, firstly, that premise (2) is fairly strong: it says that we are sometimes *obligated* to believe a proposition that is equivalent to – or trivially entails – the claim that it is *impermissible* to believe that very proposition. If we were to replace (2) with the claim that we are *permitted* to believe self-undermining propositions like ϕ, then the Self-Undermining argument would not succeed. And if the self-undermining proposition ϕ were to be replaced with a weaker proposition which said that we are *permitted* to not believe that same proposition, then the Self-Undermining argument would not succeed. So accepting premises (1), (3) and (4) of the Self-Undermining argument can at best challenge a very strong version of fallibilism: it does not challenge fallibilists who believe that we can have permissible but not obligatory beliefs in propositions like t, or that we can have obligatory beliefs in propositions that are slightly weaker than t. It therefore does not challenge the fallibilist who says that if agent's situation includes testimony that x, then the agent is nonetheless rationally permitted to fail to believe that x. This still seems like this leaves us with a fairly robust form of fallibilism. Denying (2) does not commit us to the view that the true epistemic principles always say you are never permitted to believe in false epistemic principles, as normative

[25] It might seem implausiblethat rejecting (4) in its current form will free us of the self-undermining problem. After all, if the Über Rule U is consistent with Anti-Akrasia (i.e. if we never have the most reasons to adopt an akratic attitude state) then U and Anti-Akrasia will never produce conflicting obligations. However, in the self-undermimining case involving the Über Rule, if the agent is obligated to believe $\bar{U}$ is correct then she is obligated not to believe $\bar{U}$ is correct. If we deny (4) then we can accept that $\neg P_S \neg(B(\bar{U}))$ and $\neg P_S \neg(\neg B(\bar{U}))$.

infallibilists believe. It merely commits us to the view that the true epistemic principles never say that we are *obligated* to believe in a proposition which entails that it is *impermissible* to believe in that very proposition. This seems to be consistent with a fairly robust form of fallibilism.[26]

Let us consider the third strategy against the Self-Undermining argument: the fallibilist can deny premise (1). In order to deny that self-undermining propositions like ϕ exist. she has to say that there is no situation in which ought to be such that if you believe in a proposition x, then you believe that it's impermissible to believe x. We simply assumed that it was possible for there to be self-undermining proposition like ϕ when discussing Titelbaum's original case, but the claim that there are self-undermining propositions is not a trivial one. After all, self-undermining propositions seem to involve a problematic kind of deontic self-reference: you can only believe this kind of proposition if you also believe that you are not permitted to believe it. This seems similar to liar propositions like 'this proposition is false', which, if true, entail their own falsehood. It is well known that languages which allow for statements that refer to themselves give rise to paradoxes of self-reference. And it is generally thought to be desirable to purge language of this kind of self-reference. Could this not be true of deontic languages as well?

The cases the involve self-undermining propositions given above (the testimony, über rule, and RI/RT cases) seem to constitute something akin to deontic paradoxes of self-reference. They all involve propositions that are either equivalent to – or trivially entail – the claim that it is impermissible to believe in those very propositions. The original self-undermining problem involved a case in which the Testimony principle said that if an agent's situation includes testimony that x, the agent is rationally permitted and obligated to believe that x. This became problematic when the agent received evidence for t, which said that if an agent's situation includes testimony that x, the agent is rationally forbidden to believe that x. We were assuming here that x stood for any proposition, and that t can be substituted for x. But it seems unlikely that we could substitute t for x if we employed a hierarchy of deontic languages in order to rule out paradoxes of self-reference.

One general response to paradoxes of self-reference involves appealing to a hierarchy of languages that do not allow for statements to refer to themselves. In a hierarchy of *deontic* languages, level 0 statements would presumably consist in propositions that contain no deontic operators: e.g. 'it is raining' or 'Amy believes it is raining'. Level 1 statements would consist in propositions that contain a single deontic operator like 'it is permissible to believe it is raining' or 'Amy believes it is permissible to believe it is raining'. And level n statements would consist in propositions that contain n-many embedded deontic operators, and so on ad infinitum.[27]

By claiming that there are hierarchies of deontic languages that do not allow for self reference within a level, we can stop the Self-Undermining argument in its tracks. If we do not allow for level

[26] It is worth noting that if the Reasons Down-based Top Down view is true, then we should not expect fallibilism to be trivially consistent with premise (2). On the Reasons Down-based Top Down view, $O_S(B(\neg P_S(A))) \rightarrow \neg O_S(A)$. So if there are propositions like like t, which entails $O_S(B(\neg P_S(t))$, then the Top Down view will be inconsistent with any principle, such as Testimony, which says that the agent ought to believe t. But even if the Top Down view can help us avoid inconsistency with Anti-Akrasia in these cases, it is still difficult to see what stable set of beliefs one could have in a self-undermining situation like the one that Titelbaum describes.

[27] I don't propose this as anything other than a toy view about how to create a hierarchy of languages for statements about the deontic status of attitudes. It is merely to show that introducing levels can block self-undermining.

n propositions to appear under the scope of a claim at the same level, then a level n proposition $O_S(B(\phi))$ is only well-formulated over propositions at level $n-1$ or below (i.e. ϕ cannot itself be equivalent to $O_S(B(\phi))$). Suppose that Testimony (T) is a principle is a level 1 proposition: in other words, it says $O_S(B(x))$ where x is a level 0 proposition. Since t is a level 1 proposition that says $O_S\neg(B(x))$, it does not fall within the scope of Testimony. Instead, it falls under the scope of a level 2 Testimony principle (T2): $O_S(B(\psi))$, where ψ is a level 1 proposition like t. But if the level 2 Testimony principle says that we are obligated to believe t in situation S, then this does not result in any self-undermining. Principle T2 requires that we believe in t, which is equivalent to the negation of T. But undermining T does not mean that we should not believe in t since, as a level 1 proposition, t does not fall under the scope of the level 1 proposition T. This means that an agent can come to believe $t/\neg T$ without violating either principle T2 or Anti-Akrasia.

However, there are at least three objections that this response to the Self-Undermining argument needs to deal with. First, we need to show that we can rule out trivial entailments across levels, or that such entailments do not give rise to self-undermining problems. Second, we need to show that the response will also succeed in cases that involve mutually undermining propositions (as in the RI/RT case above). Third, the response given above commits us to a hierarchy of deontic languages that some may find problematic. Rather than there being a finite set of principles that map all attitudesonto deontic operators, there is an infinite hierarchy of principles that each map attitudes of the level below themselves onto deontic operators.[28]

I won't try to tackle any the problems for this response here. My intention in pointing out that self-undermining propositions seem to involve a problematic kind of self-reference, and in gesturing towards the hierarchical view of deontic languages as a potential solution, has merely been to show that premise (1) is far from invulnerable. We may want to reject the claim that there are any well-formulated propositions that are deontically self-undermining.

The Self-Undermining argument shows that we cannot jointly accept premises (1) - (4). This is a problem for fallibilists who believe that we can, in some situations, be obligated to believe in the deontic inverse of true principles about epistemic obligations, and that we can be obligated to do so on the basis of those very principles. The normative infallibilist avoids the Self-Undermining argument by maintaining that the true epistemic obligations never permit these sorts of mistaken beliefs about themselves. But in this section I have argued that a denial of premise (2) does not force us to adopt anything as strong as normative infallibilism. Weaker forms of fallibilism can hold that we can have less radically mistaken beliefs about the principles of rationality without falling prey to the self-undermining problem. I have also argued that we can reject the claim that epistemic obligations can apply to beliefs about themselves. If we believe that either of these two responses to the Self-Undermining argument is more plausible than the view that we can never be permissibly mistaken about what we are obligated to believe, then the Self-Undermining argument does not constitute a successful challenge to fallibilism.

[28] Appealing to a hierarchy of deontic languages is not the only solution to deontic paradoxes of self-reference. We could, for example, appeal to a single deontic language in which deontic self-reference is not well formed. It seems like many responses to standard problems of self-reference could be appealed to as a solution to deontic self-reference.