Higher-order evidence and the limits of defeat

Maria Lasonen-Aarnio

This is the pre-peer reviewed version of the following article: Lasonen-Aarnio, M. (2013). Higher-order Evidence and the Limits of Defeat. Forthcoming in *Philosophy and Phenomenological Review*

Recent authors have drawn attention to a new kind of defeating evidence commonly referred to as *higher-order evidence*. Such evidence works by inducing doubts that one's doxastic state is the result of a flawed process – for instance, a process brought about by a reason-distorting drug. I argue that accommodating defeat by higher-order evidence requires a *two-tiered theory of justification*, and that the phenomenon gives rise to a puzzle. The puzzle is that at least in some situations involving higher-order defeaters the correct epistemic rules issue conflicting recommendations. For instance, a subject ought to believe *p*, but she ought also to suspend judgment in *p*. I discuss three responses. The first resists the puzzle by arguing that there is only one correct epistemic rule, an Über-rule. The second accepts that there are genuine epistemic dilemmas. The third appeals to a hierarchy or ordering of correct epistemic rules. I spell out problems for all of these responses. I conclude that the right lesson to draw from the puzzle is that a state can be epistemically rational or justified even if one has what looks to be strong evidence to think that it is not. As such, the considerations put forth constitute a non question-begging argument for a kind of externalism.

I The defeatist trend in epistemology

There has been an ever-accelerating trend in post-Cartesian epistemology to acknowledge the defeasibility of justification and knowledge. The scope of beliefs regarded as defeasible has broadened because of the recognition of new kinds of evidence with putative defeating force. The newest addition to this category is a variety of evidence that has been dubbed "higher-order" or "second-order" evidence.[1] In one way or another, such evidence works by inducing doubts that one's doxastic state is the result of a flawed process. In what follows I will argue that attempts to take defeat by higher-order evidence seriously face a puzzle. I sketch three responses to the puzzle, and argue that there is no altogether happy resolution.[2] But let me begin by saying a bit more about the putative phenomenon of defeat by higher-order evidence.

Evidence that one's doxastic state is the result of a flawed process can take many forms. It may be evidence that one is subject to a deep but undetectable cognitive malfunction; that one has made a simple calculation error; that one has failed to appreciate the import of one's evidence; or even that the epistemic rules one follows are

[1] See, for instance, Christensen (2007a, 2007b, 2007c, 2009, 2010a), Elga (unpublished), Feldman (2005), Kelly (2010), Schechter (2011).

[2] I see no significant difference between intuitions elicited by more familiar cases of defeat and those elicited by cases involving higher-order evidence. This is why I take the kinds of problems sketched below to cast doubt on a defeatist way of thinking more generally.

incorrect. Here are some examples involving the kind of higher-order evidence that is widely thought to have defeating force:

> ***Mental maths***
> My friend and I have often amused ourselves by solving little math problems in our heads, and comparing our answers. We have strikingly similar track records: we are both very reliable at doing mental maths, and neither is more reliable than the other. We now engage in this pastime, and I come up with an answer to a problem, 457. I then learn that my friend came up with a different answer, 459.[3]

> ***Hypoxia***
> I have just achieved a difficult first ascent in the Himalayas. As the weather turns, I have to abseil down a long pitch. I have gone through a sequence of reasoning several times to check that I have constructed my anchor correctly, that I haven't under-estimated the length of the pitch, and that I have threaded the rope correctly through my belay device and carabiner. I then acquire evidence that I am in serious danger of being affected by a mild case of hypoxia caused by high altitude. Such hypoxia impairs one's reasoning while making it seem perfectly fine. I know that mountaineers have made stupid but fatal mistakes in the past as a result of being in such a condition.[4]

In both cases I acquire evidence that I am subject to a cognitive malfunction of some sort and hence, that my doxastic state is the output of a flawed cognitive process.[5] Though this is not uncontroversial, I will assume that the kind of flaw we are interested in is one that would prevent a belief from being epistemically rational or justified in the first place: if I make a simple calculation error or reason badly due to hypoxia, my belief may be excusable, but it is not justified. Hence, the kind of rational self-doubt induced by higher-order defeaters is doubt about one's own epistemic rationality. By contrast, evidence that a belief is false, for instance, need not be evidence that it is flawed in this sense.[6]

Among current epistemologists the majority view appears to be that evidence producing sufficiently strong rational doubt about the epistemic integrity of a doxastic state calls for revising that state, even if it is the result of an impeccable process.[7] These epistemologists endorse a principle along the following lines:

---

[3] The case closely resembles cases often used to motivate conciliatory views of disagreement. See, for instance, Christensen (2010a: 186-187).

[4] Cf. Christensen (2010b), Elga (unpublished) and Schechter (2011).

[5] What I will mean by "acquiring evidence that I am subject to a cognitive malfunction" or "acquiring evidence that my doxastic state is the output of a flawed cognitive process" is acquiring evidence that now makes it sufficiently likely – likely above some threshold – that such an eventuality has occurred. The threshold need not be that required for outright belief. For instance, if in *Mental maths* I am equally confident that my friend has made a mistake as that I have, but confident that not both of us are mistaken, then I am 50% confident that I am subject to a cognitive malfunction. Presumably, this doesn't justify outright belief.

[6] What about evidence that a belief isn't reliably formed? On the present approach whether or not such evidence always has higher-order defeating force depends on whether reliability is a necessary condition on justification. See also footnote 8.

[7] But there are exceptions – see, for instance, Field (2000).

Higher-order defeat

Evidence that a cognitive process producing a doxastic state S as output is flawed has defeating force with respect to S.[8]

Coarse-grained states such as believing a proposition *p* or suspending judgment in *p* count as doxastic states, but so do more fine-grained states such as assigning a credence of 0.8 to *p*. So, for instance, my 0.8 confidence in rain tomorrow may no longer be justified or rational if I acquire evidence that I have misevaluated the meteorological data. *Higher-order defeat* yields a simple recipe for generating a defeater for just about any belief: just imagine sufficiently strong evidence that the cognitive process producing that belief is flawed. I will simply take on board the view that in cases of the sort described above the subject does, indeed, acquire evidence that some of her cognitive processes are flawed. The question will be whether such evidence has defeating force with respect to the states that are outputs of those processes.

What has been said suggests an account of how higher-order defeat differs from more ordinary or traditional kinds of defeat. Consider a case in which I come to believe some proposition *p*, but my belief is defeated by subsequently acquired evidence for not-*p*. Such rebutting evidence need not be evidence that my original belief in *p* was flawed in a way that prevented it from ever being justified. Indeed, can't I recognise that my original belief was perfectly justified given the evidence I had? The same applies to standard examples of undercutting defeat. Assume that I am told that what I believe based on perception to be a red object is in fact being illuminated by red trick lighting. Such evidence doesn't seem to cast any doubt on the epistemic rationality of my originally believing the object to be red based on my perceptual experience as of a red object. By contrast, defeat by higher-order evidence has a retrospective aspect, providing a subject with evidence that her belief was never rational, reasonable, or justified to start out with.[9] It's not just that now, once I get evidence that I may be suffering from hypoxia, I am no longer justified in believing what I did. I acquire evidence that I was never justified to start out with. I will assume that whether or not it suffices to distinguish

---

[8] The principle is similar in spirit to a principle Christensen calls *Integration*:

> ***Integration***
> An agent's object-level beliefs must reflect the agent's meta-level beliefs about the reliability of the cognitive processes underlying her object-level beliefs (Christensen 2007b).

The reason why I prefer not to speak of evidence about the reliability of one's cognitive processes is that I doubt whether having such evidence is sufficient or necessary for the kind of phenomenon that people tend to classify as defeat by higher-order evidence. Imagine that I acquire evidence that I am a brain in a vat. Wouldn't this be evidence that processes underlying my object-level beliefs about my environment fail to be reliable? However, it is unclear whether such defeaters should be classified as higher-order, for it is at least controversial whether I acquire evidence that my states are flawed in a way that prevented them from ever being epistemically rational or justified. After all, many think that it is rational for a brain in a vat who is unaware of her predicament to take perceptual experiences at face value. Also, if a reliabilist account of justification is rejected, it should be possible to acquire evidence that a cognitive process fails to produce a rational belief without acquiring evidence bearing on the reliability of the process.

[9] This is all compatible with a piece of evidence having both higher-order and first-order defeating force. Perhaps, for instance, higher-order defeaters often or even always have some rebutting force: evidence that I came to believe *p* as a result of a flawed process may be weak evidence that *p* is false.

ordinary from higher-order defeaters, higher-order defeat has the kind of retrospective aspect described.

Now, someone might grant that there is a difference between higher-order defeaters and more ordinary kinds of defeaters, but wonder whether there is really anything that puzzling about defeat by higher-order evidence. Consider the following line of thought:

> "Epistemic rationality is a matter of proportioning one's doxastic states to the evidence. For instance, one ought to believe *p* just in case one's evidence sufficiently supports *p*. But when a subject has higher-order evidence that a doxastic state she is in is flawed, retaining that state doesn't count as proportioning her beliefs to her overall evidence. Higher-order evidence is just more evidence. There is nothing puzzling about it"

The suggestion is that higher-order evidence that one's belief in *p* is flawed affects the degree of support that one's total evidence lends to *p*, for such evidence is ipso facto evidence against *p*. I am very sympathetic to the thought that higher-order evidence should be treated as just more evidence. However, whether or not this will yield anything like the systematic sort of defeat that a lot of epistemologists are after – that is, a view respecting a principle along the lines of *Higher-order defeat* – is far from clear.

First, one might wonder whether higher-order evidence has the desired effect on degrees of support. In effect, David Christensen argues that it is peculiar to such evidence that conditionalising on it leaves the degree to which one's evidence supports the relevant proposition *p* intact.[10] Whether or not this is right, I think the above line of thought runs into problems not deriving from any specific norms for belief-revision or credal dynamics. Assume for the sake of argument that conditionalising on higher-order evidence had the desired effect on one's credences, and that conditionalisation was in fact *the* correct way of taking new evidence into account. Now consider evidence that one has been given a drug that makes everything seem perfectly normal, though one is in fact highly likely to make mistakes in attempting to conditionalise on evidence. If new evidence ought to be taken into account by conditionalisation, one ought to conditionalise on this (and subsequent) evidence. But *Higher-order defeat* would suggest that the resulting doxastic states are defeated, for one would have evidence that they don't result from conditionalising on one's evidence. And so it looks like there is evidence that cannot be rationally taken into account by conditionalisation after all.

This points to the kind of puzzle created by attempting to take systematic defeat by higher-order evidence seriously. In what follows I want to spell out the puzzle in more detail, and to discuss different resolutions. Ultimately, I think that each runs into problems that warrant re-evaluating principles like *Higher-order defeat*. I will begin by arguing that endorsing defeat by higher-order evidence creates pressure to adopt what I call a *two-tiered theory of justification.*

[10] Christensen (2010a: 195).

II A two-tiered theory of justification

A two-tiered theory of justification is generated by the following thought. Consider candidates for the kind of epistemic good **F** that confers justification (or epistemic rationality) on a doxastic state. **F** might, for instance, be the property of having come about by an application of correct epistemic rules, or the property of being the output of a reliable process. Now assume that it is possible for a state to have **F** while one has evidence that it lacks **F** – or, more generally, that it is possible for a state to have **F** while one has evidence that the state in question is flawed. Then, **F** cannot be a necessary and sufficient condition for justification or epistemic rationality after all. An additional condition is needed that rules out higher-order defeaters. A useful heuristic picture of what a two-tiered theory of justification will involve is the following: if one's belief in a proposition *p* is to be justified, it must have some property **F**, and in addition, one must lack (sufficiently strong) evidence that it fails to have **F**.

In what follows, I will make an assumption about what the epistemic good-making properties of doxastic states are. It has its roots in what I will call a *rule-driven picture of epistemic cognition*. On such a picture, doxastic responses at least typically involve the application of epistemic rules, and whether or not a doxastic state is epistemically rational or justified depends on the goodness of the rules that were applied in forming or maintaining that state. I will call the good rules the *correct* ones, recognising that in the two-tiered theory I sketch, correct rules play a somewhat different role than they do in more traditional theories exemplifying the rule-driven picture. The idea is roughly (subject to some qualifications below) that it is a necessary condition on the rationality of a doxastic response that it involve the application of a correct epistemic rule. Those who endorse the picture tend to think that this is also a sufficient condition. But for reasons that will become clear, the phenomenon of higher-order defeat puts serious pressure to reject any assumption of sufficiency.

The rule-driven picture will provide a neat framework for exploring what is puzzling about higher-order evidence. Though not uncontroversial, it is very widely accepted.[11] In particular, the most well-developed and promising attempts to come to terms with more ordinary kinds of defeaters subscribe to, or are at least compatible with, thinking about epistemic rationality in terms of following correct rules or norms.[12] The picture is able to accommodate a very wide range of views. For instance, any view interested in a notion of well-founded belief, belief that is based in the right sort of way on adequate and undefeated reasons or evidence, will presumably need to invoke the idea of correct rules of inference in saying what it is to base a belief on the right reasons in the right kind of way. Or, take a view on which justified or epistemically rational degrees of belief obey the probability axioms and evolve by some form of conditionalisation. Presumably, proponents of such views think of a rule telling one to conditionalise on new evidence as a correct epistemic rule.

I will assume that an epistemic rule can be represented by a function from possible circumstances to recommendations about which doxastic states one may or must adopt in those circumstances. But there is controversy about exactly how we should think

[11] Cf. Boghossian (2008). Though Boghossian raises problems for the picture, he admits to finding it "as natural and compelling as the next person" (473).

[12] For instance, Pollock and Cruz (1999) write: "a belief is justified if and only if it is licensed by correct epistemic norms"(123).

of rules in general, and epistemic rules in particular. On one view, rules are imperatival contents along the following lines:

> In circumstances of type C, believe *p*!

On another, they are normative propositions along the following lines:

> In circumstances of type C, one ought to/is permitted to believe *p*.

I won't need to adjudicate between these.[13] Even on the first view, I take correct epistemic rules to be associated with true statements about what one ought to or may believe in various circumstances – indeed, I take this to follow from the fact that they are normative. On some of the views to be discussed these will emerge to be merely *prima facie* oughts. But this need not worry us just yet.

Statements along the lines of "In circumstances of type C, one ought to believe *p*" are ambiguous, for the "ought" in such statements can be assigned either a scope over the whole conditional or just its consequent. I doubt that much of what I say below rests on this choice. However, considering cases of higher-order defeat may give us a reason to take them to be wide-scope. Take a recommendation along the following lines: "If you believe *p* and acquire such-and-such evidence that your belief is flawed, then revise it in such-and-such a way". If we construe the normative proposition corresponding to this statement as involving a narrow-scope ought, then it looks like the rule would only ever apply to subjects who believe *p* while having evidence that their beliefs are flawed. But if champions of defeat by higher-order evidence are right, it is epistemically irrational to believe *p* while having evidence that one's belief is flawed. A rule that can only ever apply to subjects who are (at least for some time) irrational seems problematic. The issues here are complex, and I don't want to rest too much on this observation. But it might provide a reason to take the epistemic rules in question to correspond to wide-scope oughts.[14]

With the rule-driven picture on the table, let me now return to the argument for a two-tiered theory of justification, and the puzzle to which defeat by higher-order evidence gives rise. Again, let **F** be a property capturing the kind of epistemic goodness that a doxastic state must exemplify in order for that state to be epistemically rational, reasonable, or justified. Taking the rule-driven picture as a starting point, I will assume that **F** is roughly (subject to some qualifications below) the property of coming about by applications of correct epistemic rules, be these rules telling one to take one's sensory perceptions (or one's memories, or the testimony of others) at face value when there is no reason to distrust them, or rules like conditionalisation.

Such a picture may look congenial when it comes to ordinary kinds of defeaters. Consider, for instance, the kinds of epistemic rules governing sensory perception that epistemologists have proposed. A rule telling one to believe *p* when it perceptually seems to one as if *p*, *no matter what*, doesn't look like a very good candidate for a correct rule.

[13] For a discussion – and problems for both views – see Boghossian (2008).

[14] Perhaps imperatives are ambiguous in a way that mirrors the ambiguity in the corresponding normative statements. Should it, however, turn out that imperatival contents correspond to the narrow-scope propositions, this would be a reason not to think of epistemic rules as imperatival contents.

A rule telling one to believe *p* when it perceptually seems to one as if *p* and there is no (sufficiently strong) reason to think that one's perceptions are not to be trusted is much better.[15] Similarly, a view on which a subject ought to conditionalise on new evidence seems able to make sense of at least many ordinary kinds of defeaters: the probability of an object being red conditional on its looking red is high, but the probability of being red conditional on looking red and being illuminated by trick red lighting is not.[16]

Be the matter of ordinary defeaters as it may, when it comes to dealing with higher-order defeaters, the property **F** would have to be one that beliefs can only have in the absence of evidence that they are flawed. In so far as a state must have **F** in order to be justified or epistemically rational, evidence that a state lacks **F** is (at least sometimes)[17] evidence that it is flawed, or fails to be epistemically rational. Then, the property would have to satisfy the following condition:

> It is impossible for a subject's doxastic state to have **F** if the subject has evidence that her state lacks **F**.

Hence, having evidence that one's state lacks **F** will entail that it in fact lacks **F**. Are there viable candidates for justification-conferring properties that satisfy this condition?[18] Consider, in particular, the following condition:

> It is impossible for a subject's doxastic state to be the result of applying correct epistemic rules if the subject has evidence that her state is not the result of applying correct epistemic rules.

The thought is that no correct epistemic rules allow being in a doxastic state S in the presence of evidence that one has failed to apply correct epistemic rules in coming to be in S. But, one might wonder, can't a belief result from applying an impeccable rule of inference even, for instance, in the presence of strong evidence that someone has messed with one's inferential abilities, or that one has fallen victim to hypoxia? I will now give some tentative reasons to think that this thought is right, and that the above condition fails. I bolster this conclusion later, in connection with the Über-rule view.

---

[15] One often sees epistemologists formulate rules broadly along these lines. Take, for instance, "If something looks red to you and you have no reason to think that it is not red then you are permitted to believe it is red" (Pollock & Cruz 1999: 157), or "Believe *p* whenever you have an experience or apparent perception as of *p*'s being the case, and have no special reason to think that your experiences are unreliable in the circumstances" (Wedgwood 2002: 276).

[16] Though of course, such a picture cannot handle seeming defeaters for doxastic states that have as their contents propositions that are entailed by one's evidence. I have serious doubts about the ability of the rule-driven picture to handle even ordinary kinds of defeaters, but this is not the place to pursue them.

[17] One might worry about subjects who lack knowledge that **F** is the correct justification-conferring property – would evidence that a state lacks **F** count as evidence that it is flawed for those subjects? However, at least for subjects who know that **F** is the justification-conferring property, evidence that a state lacks **F** would count as evidence that it is flawed. Such subjects are clearly possible.

[18] Of course one can build such a property by brute force: let **F*** be the property a state has just in case one doesn't have evidence that it lacks **F***. To avoid this dubiously circular structure, one could instead construct **F*** as the property a state has just in case there is no property **G** (such that **G** ≠ **F***) such that one has evidence that one's state lacks **G**. It is unclear whether a state would or even could ever have **F***. But more importantly, such properties are hardly candidates to confer justification on doxastic states.

First, I am not at all convinced that the epistemic rules we follow and regard as correct have provisos for higher-order defeat built into them. Take, for instance, a rule telling one to believe *p* when it perceptually seems to one as if *p* and there is no reason to think that in the present situation, how things seem isn't a good guide to how they are. Call this rule *Perception*. But now consider a situation in which I believe that it is snowing outside based on its perceptually seeming to me as if it is snowing, but I also have strong but misleading evidence that I have been given a drug that seriously impairs my ability to form beliefs based on following *Perception*: the drug makes me prone to misjudge how things seem to me. This is not evidence that, in my present circumstances, how things seem is not a good guide to how they are. Moreover, because the higher-order evidence is misleading, it does in fact perceptually seem to me as if it is snowing. Hence, in my circumstances *Perception* still recommends believing *p*. Or, take a rule urging one to infer *q* from *p* and *if p, q* in situations in which one knows (or justifiably believes) both *p* and *if p, q*. The rule urges one to infer *q* from *p* and *if p, q* even in situations involving evidence that *modus ponens* is not a valid rule of logic. In so far as any such evidence is misleading, isn't the unqualified inference rule perfectly correct?

It is worth noting in passing that the idea that correct epistemic rules can continue to apply in the presence of higher-order defeaters would provide a diagnosis of a thought defended in several places by David Christensen, namely, that in situations involving higher-order defeaters, subjects are forced into a kind of epistemic imperfection:

> Even though she's capable of perfect logical insight, and even if she flawlessly appreciates which hypotheses best explain the evidence she has, she cannot form the beliefs best supported by those logical or explanatory relations that she fully grasps. So…in taking HOE [higher-order evidence] into account in certain cases, an agent is forced to embody a kind of epistemic imperfection.[19]

Pointing out that correct epistemic rules can still recommend belief in the presence of higher-order defeaters would be a step towards explaining such intuitions: the imperfection in question would result from failing to follow the recommendations made by perfectly correct inductive and deductive rules.

Now, someone might object that the kinds of rules I have described are not correct. Instead of the rule *Perception*, for instance, consider a rule *Perception**, which urges one to follow *Perception* with exceptions: one ought not to follow *Perception* in situations involving evidence of the kind described above, evidence that one is prone to misapply *Perception*, or evidence that *Perception* is incorrect. If one thus specifies the right epistemic rules to start out with, one might wonder, is there any need for an additional condition on justification and hence, what I am calling a two-tiered theory? But of course, a subject can acquire evidence that she has misapplied *Perception**, or that *Perception** is incorrect. Though *Perception** can prevent a subject from inferring by *Perception* in situations involving higher-order defeaters of the relevant sort, it cannot prevent one from inferring by *Perception** itself. I will leave discussion of an Über-rule that by definition never encounters such problems for later.

I have given two tentative reasons to think that following correct epistemic rules cannot be a sufficient condition on epistemic rationality. First, having built-in provisos

[19] Christensen (2010a: 193)

for higher-order defeaters just doesn't seem like a criterion for being a correct epistemic rule. And second, even when a rule has some such provisos built into it, it need not have provisos for *all possible* higher-order defeaters built into to it – and it is at least a *prima facie* challenge to show how this would even be possible. More generally, unless the justification-conferring property **F** is a property that a doxastic state can have only if one lacks evidence that the process producing the state is flawed in a defeat-inducing way, a theory of justification able to accommodate defeat by higher-order evidence must admit an additional condition – a condition stating that the subject must lack evidence of the defeating kind. And so we have arrived at a *two-tiered theory of justification*. Such a theory has roughly the following form: A doxastic state S is justified just in case (i) S has **F**, and (ii) one lacks evidence that S is flawed. In so far as what it is for a state to be flawed is just for it to fail to have **F**, one might hope to state (ii) neatly as follows: the subject must lack evidence that her state lacks **F**.[20]

Within the context of a rule-driven picture, the rough idea will be that a doxastic state is justified just in case (i) it is the product of following a correct epistemic rule, and (ii) one doesn't have evidence that it is flawed. A typical example of evidence that one's state is flawed is evidence that it is the result of a failure to apply correct epistemic rules. Such evidence can take various forms. It could be evidence that one has failed in one's attempt to apply what are in fact perfectly correct epistemic rules, it could be evidence that the rules one has successfully applied are faulty, or it could be evidence that favours neither of these possibilities. Compare: one might acquire evidence that a device fails to run a program, evidence that it runs a faulty program, or evidence that is ambiguous between these alternatives. Evidence that a rule one has applied is incorrect is particularly far-reaching, for it threatens to infect all states that result from applications of the rule. Indeed, such evidence will play an important role in the discussion below.

I have given a rough statement of the two-tiered theory I want to focus on, but a bit more fine-tuning may be needed. On one picture, all rational doxastic responses involve applications of correct epistemic rules. But there is also a more economical picture on which a doxastic response is rational if it is either an application of a correct epistemic rule, or else it is not prohibited by the correct rules. For instance, assume that I correctly suspend judgment about whether I ate caramel cake on my third birthday in circumstances in which I seem to remember that I did, but a reliable testifier says otherwise. The testifier then admits to not having a clear memory of the relevant event after all. But assume that the correct rules don't say what I must do in the new circumstances: there is no rule prohibiting me from continuing to suspend judgment on the matter, and there is no rule prohibiting me from believing that I did indeed eat caramel cake on my third birthday. The thought, then, is that it is rational for me to do either. But it might seem strained to say that if I come to now form the relevant belief about my third birthday, I am following a correct epistemic rule.[21] Hence, proponents of a rule-driven picture need not be committed even to the idea that it is a necessary condition

[20] However, there are various issues that arise here. First, one might wonder about subjects who don't know that **F** is the justification-conferring property, but acquire evidence that their state lacks **F**. And second, what about subjects who have a false theory to the effect that the justification-conferring property is some property **G**, and who acquire evidence that their state lacks that property? In so far as such a false theory is sufficiently well supported, wouldn't evidence that one's state lacks **G** count as evidence that it is flawed?

[21] This is so even if there is a rule that says that in my present circumstances I shouldn't believe that I *didn't* eat caramel cake on my third birthday.

on the rationality of any doxastic state that it has come about through the application of a correct epistemic rule.

With this qualification on the table, the two-tiered theory of justification I wish to focus on can be formulated as follows:

> **Condition 1**
> A doxastic state S is epistemically rational only if [*either*] it is the result of following correct epistemic rules [*or it is at least not prohibited by the correct rules*].
>
> **Condition 2**
> A doxastic state S is epistemically rational only if one lacks evidence that it is flawed.[22]

I am reluctant to state Condition 2 as the condition that one lacks evidence that one's state fails Condition 1, for even subjects who lack knowledge of what the correct justification-conferring property is might acquire evidence that their states are flawed in a way that prevents them from being justified. However, as noted above, evidence that a state did not come about by applying correct epistemic rules certainly seems like a paradigm form of evidence that the state is flawed.

Though I focus on theories that endorse the rule-driven picture, I do not mean to imply that views that fall outside it can avoid a two-tiered structure. Take, for instance, a reliabilist theory on which a belief is justified or epistemically rational just in case it is the result of a reliable process (and forming beliefs by reliable processes doesn't amount to following correct rules). The problem for such a theory is created by the observation that it would seem possible for a belief to be the result of a perfectly reliable process even in the presence of evidence to think that it is not. Imagine, for instance, that I possess an infallible logical faculty, and have come to believe *p* based on a deliverance of this faculty. I then acquire evidence that I have been given a pill that causes me to make bad, undetectable mistakes in my reasoning. Couldn't I continue to believe *p* as the result of a process that is 100% reliable in such circumstances? Reliabilists have made various moves to avoid an explicitly two-tiered structure, none of which I find convincing. For instance, one could suggest that justified belief in *p* requires not only that one's belief be the result of a reliable process, but that there be no alternative, equally reliable processes available to one that would not result in a belief in *p*.[23] But it looks like no alternative process by which I might come to give up belief in *p* would be 100% reliable. And besides, the proposed view would not yield the result that the subject would be justified in giving up her belief in *p*, since there would be an alternative, competing process by which she could continue believing *p*.

I have made a provisional case for the claim that a theory on which following correct epistemic rules is both a necessary and sufficient condition for epistemic rationality or justification is unable to accommodate the phenomenon of defeat by higher-order evidence – and hence, that a two-tiered theory of justification is needed. I will spell

[22] Recall that by "evidence that a state S is flawed" I mean evidence that brings the likelihood that it is flawed above some threshold.
[23] See Goldman (1986).

out a puzzle created by the two-tiered theory, and consider three responses. Before doing so, however, I want to consider an argument to the effect that adding Condition 2 still doesn't take us even close to a set of sufficient conditions on justification.

I have assumed that higher-order defeat only ever happens as the result of evidence that one's doxastic state is flawed, where being flawed just is lacking the justification-conferring property **F**. But one might worry that this isn't the only form that defeat-inducing evidence can take. For what about evidence that the component of justification formulated by Condition 2 fails to be satisfied that is *not* evidence that Condition 1 fails to be satisfied? This would nevertheless be evidence that one's belief fails to be justified or epistemically rational, since Condition 2 is a necessary condition on justification. But isn't evidence that one's belief fails to be justified or epistemically rational higher-order evidence with defeating force with respect to that belief? Those convinced by this could hope to remedy the problem by altering Condition 2 as follows:

> **Condition 2***
> A doxastic state S is epistemically rational only if one lacks evidence that S fails to be epistemically rational.

Condition 2* would entail, then, that one of the ways in which a belief can be defeated is if one has evidence that it fails to satisfy Condition 2* itself. Condition 2* is circular in a manner that may seem problematic. For this reason, proponents of the idea that justification or epistemic rationality entails lacking evidence that one is not justified might prefer to opt for a theory that has infinitely many conditions, taking something like the following form:

> Condition 1: X
> Condition 2: One lacks evidence that one's state is flawed.
> Condition 3: One lacks evidence that one has evidence that one's state is flawed.
> Condition 4: One lacks evidence that one has evidence that one has evidence that one's state is flawed.
> .
> .
> .

In so far as evidence that one's belief is not justified is always evidence that one of the conditions fails to be met, such a theory would yield the result that evidence that a belief is not justified entails that it is not justified.[24] So perhaps only a theory with infinitely many tiers could fully accommodate the phenomenon of higher-order defeat.

Everything I say below is compatible with this theory being right. The sort of puzzle created by a two-tier structure is also created by a structure with infinitely many

[24] Cases in which a subject has evidence that one of the conditions fails to be met, without having evidence, of any one of these conditions, that it is met, are trickier. For such cases certainly seem like ones in which a subject has evidence that her belief fails to be justified (since she has evidence that not all conditions on justification are met) without needing to fail any of the conditions on justification. Such cases might show that there is no way of avoiding a circular condition along the lines of Condition 2*. But the remarks made below apply equally to Condition 2*.

tiers (or a theory that replaces Condition 2 by Condition 2*). However, I am not convinced that even the most astute proponents of higher-order defeat are pressed to adopt it. In fact, I think that there are excellent reasons not to. For consider the kind of evidence that, on this account, would have defeating force, but that Condition 2 would not suffice to accommodate. It might be evidence that one has evidence that one's state is flawed (perhaps, evidence that one has evidence that one has failed to follow correct rules), without being evidence that it is flawed (evidence that one has failed to follow correct rules). Indeed, on the proposed theory, *any* number of such iterations, no matter how great, would entail lacking justification. Now, sometimes one hears the slogan "evidence about evidence is evidence". In so far as the thought is that evidence that one has evidence that *p* is entails having evidence that *p*, I think the slogan is simply mistaken. Its being probable that it is probable that *p* doesn't entail that it is probable that *p*. But in any case, if the slogan were true, there would be no need for infinitely many tiers in the first place, for failing any Condition 2 + *i* would entail failing Condition 2. The need only arose from recognizing the falsity of the slogan. But in that case one wonders why, for instance, evidence that one has evidence that one has evidence that one has failed to follow correct epistemic rules isn't too far removed to have any defeating force. Such evidence may fail to make it sufficiently likely that one's belief is flawed to have any defeating force.

Below I will say more about theories that may be able to avoid a two-tier structure. But I think there is at least a *prima facie* case to be made in its favour. Let me now finally state the kind of puzzle created by the two-tiered theory of justification sketched.

III A puzzle

Assume that a correct epistemic rule R recommends (requires) believing *p* in circumstances C, and that Suzy is in fact in circumstances C. Suzy comes to believe *p* as a result of applying R. But she then acquires evidence that her doxastic state was not the result of applying a correct rule and is, therefore, flawed (this may or may not be evidence that R itself is flawed). Assume that despite possessing this higher-order evidence, Suzy continues to be in circumstances in which R recommends believing *p*.[25] Nevertheless, by Condition 2, believing *p* is no longer epistemically rational. However, those who defend defeat by higher-order evidence typically claim that at least in some such cases, there is a rational way (or a range of rational ways) of revising one's doxastic state.[26] Perhaps Suzy ought now to suspend judgment in *p*. But if suspending judgment in *p* is to be epistemically rational, then the state of suspending judgment in *p* must itself satisfy Condition 1.

Assume first a view on which all rational doxastic responses are the results of applying correct epistemic rules. Then, there must be a correct epistemic rule telling Suzy to suspend judgment in *p*. But since circumstances C still obtain, there is also a correct epistemic rule, rule R, still telling Suzy to believe *p*. The upshot is that there is a correct epistemic rule urging Suzy to believe *p* in her present circumstances, and there is a

[25] Indeed, it was the possibility of such cases that created the need for a two-tiered theory of justification.
[26] Note that Condition 2 alone does not guarantee this: violating it entails that a subject's doxastic state fails to be epistemically rational, but this does not entail that some specific way of revising it would be rational.

correct rule urging her not to believe *p* in her present circumstances. If correct rules are accompanied by oughts, it looks like Suzy ought to believe *p*, and she ought to suspend judgment in *p*. Exactly how should we conceive of the structure of correct rules that make incompatible recommendations, and exactly how should we think of the normative force of these rules?

What about the alternative, more economical view mentioned above on which being produced by applying correct rules isn't a necessary condition on the rationality of a doxastic state? Unfortunately, the rationality of suspending judgment in *p* cannot be explained by appeal to the idea that sometimes a state can be epistemically rational just because it isn't prohibited by any of the correct rules – for there *is* a correct epistemic rule prohibiting Suzy from suspending judgment in *p*, namely, the rule R she originally applied in coming to believe *p*. If it is now rational for Suzy to suspend judgment in *p*, there must be some rule urging Suzy to give up her belief in *p* (or there must at least be correct rules that don't allow Suzy to continue believing *p*). The same puzzle arises: the correct rules seem to issue incompatible recommendations.

The puzzle cannot be dissolved simply by maintaining that correct epistemic rules should be formulated in terms of what one is *permitted* or not permitted to do, not in terms of what one *must* do. It is true that there is nothing puzzling about being permitted to believe *p* while equally being permitted not to believe *p*. But rules formulated in terms of permissions can – and had better – have entailments for what one ought to or must do. For instance, if I am staring outside in ideal perceptual conditions, and can see that it is not raining, it is plausible that I ought to believe that it is not raining. I am simply not permitted to suspend judgment on the matter. And in particular, those concerned with being able to accommodate defeat by higher-order evidence won't settle for the result that subjects are merely permitted to adjust their doxastic states in the presence of higher-order defeaters: they are positively required to do so.

First assume, as before, that there is a correct epistemic rule R that requires Suzy to believe *p* in circumstances C – or at least that there is a set of rules that collectively require Suzy to believe *p* in circumstances C. Suzy then acquires the relevant sort of higher-order evidence, but she continues to be in circumstances C. By R (or the set of rules in question), she is still required to believe *p*. By rules governing higher-order defeat, she is not permitted to believe *p*, but is instead required to be in some other doxastic state, even if there is no unique alternative state she is required to be in. So there are correct rules both telling Suzy to believe *p* and telling her not to believe *p*. Again, how should we think of an epistemic system containing rules that make such conflicting recommendations? Finally, even if R was a rule not requiring, but merely permitting, Suzy to believe *p* in circumstances R, it is far from obvious that the puzzle would disappear. For then there would be a correct rule (or correct rules) telling Suzy that she may believe *p*, and there would be correct rules telling Suzy that she may not believe *p*. But she cannot both be permitted to believe *p* and not be permitted to believe *p*.

Below I discuss three responses to the puzzle. One response simply endorses it, conceding that cases of defeat by higher-order evidence create epistemic dilemmas. I argue that friends of higher-order defeat should be very unhappy with such a diagnosis. The other two paint broad, alternative pictures of how we should think about correct epistemic systems. The first rejects the two-tiered theory of justification sketched above – in particular, the thought that a belief can be formed by a perfectly correct rule, while still

being defeated. Instead, it posits an Über-rule that makes recommendations about which doxastic responses are rational in any possible situation (at least any situation in which there is some rational response), recommendations that cannot, by the very nature of the Über-rule, be defeated. The second structures epistemic rules in a hierarchical fashion, allowing for one correct rule to overrule another. I will say why I find neither of these pictures appealing. But I don't pretend to give a conclusive argument against either. Instead, my primary aim is to map out the options that remain for those who endorse something along the lines of *Higher-order defeat*. Those, like myself, who are unhappy to endorse either option should take the above puzzle as a reason to re-evaluate whether we should even try to accommodate systematic defeat by higher-order evidence.

Let me begin with the Über-rule view.

IV Über-rules

Above I claimed that correct epistemic rules need not and don't make provisions for all possible higher-order defeaters. But maybe I have failed to consider the kinds of epistemic rules that are ultimately correct. Someone might object: "If there are situations in which it is no longer rational to apply a rule R, then this shows that it was never a correct rule to start out with. And if there are situations in which it is no longer rational to apply a rule R* that tells one to infer by R *except* in certain situations involving higher-order defeaters, then that shows that R* was never a correct rule to start out with". The thought is that for any possible situation – at least any situation governed by epistemic norms – there is a rational doxastic response, and one can encode these rational responses in an overarching "Über-rule"[27]. The rule is identified by a function from epistemic circumstances to whatever the correct doxastic response is (or whatever the permitted doxastic responses are) in those circumstances.

The Über-rule view solves the puzzle described above by not allowing it to arise: because there is only one correct epistemic rule, and it never issues incompatible recommendations, situations in which correct epistemic rules make incompatible recommendations are impossible. I take the following, then, to be definitional of what an Über-rule is. First, the rule is complete in the following sense: for any circumstances in which there is some epistemically rational doxastic state in the first place, the Über-rule codifies what that state (or range of states) is. And second, though we are assuming that higher-order defeat is a real phenomenon, the following kind of situation can never arise: a subject does exactly as the rule recommends, but she has evidence that the resulting doxastic state is flawed.

But is an Über-rule even possible? Finding a rule not susceptible to defeat is surely harder than merely defining one to be such! Consider, for instance, the possibility of evidence that the Über-rule makes incorrect recommendations in some circumstances, where those include the circumstances a subject is in upon acquiring that very evidence. To have a more concrete case in mind, assume that you are staring at a chart representing the Über-rule: for each possible epistemic situation (or each relevant type of situation), the chart specifies what the recommendations made by the Über-rule in that situation are. (Let us set aside worries having to do with there being infinitely many such situations.) Now imagine that you hear an epistemology oracle tell you that the recommendations

[27] I borrow the term from Christensen (2010a).

made by the Über-rule in the very situation you are in *right now* are incorrect. In so far as the rule is complete in the sense specified above, the chart must say something about your current situation. Imagine that, as the chart tells you, the rule recommends being in state S. But in so far as the oracle is to be trusted, doesn't her testimony act as a higher-order defeater for any such recommendation?

But perhaps this doesn't show that an Über-rule is impossible after all. If nothing else, it is open to its proponents to allow the function corresponding to the rule to be undefined in certain circumstances. Here is a heuristic picture. Take a candidate Über-rule, and go over every recommendation it makes in every possible situation. Some of these may be ones in which the rule recommends being in a doxastic state, but in which one has evidence that the state in question is flawed. Now, either change the recommendation made by the rule into one that is not defeated, or – if that doesn't work – simply leave the rule undefined. The circumstances in which the rule is undefined are ones that fall outside the purview of epistemic rationality.[28] In an attempt to give the view a run for its money, I will grant that it is possible to thus pick and choose a function that yields a rule with the desired features and hence, that rules with the characteristic features of Über-rules are at least possible.[29]

It is worth emphasizing how utterly different an Über-rule is from the kinds of epistemic rules we ordinarily take ourselves to follow, or that both ordinary folk and epistemologists take to be correct. In so far as our ordinary epistemic rules don't make provisions for all possible higher-order defeaters, too bad for them: on the Über-rule view this just shows that they are incorrect to start out with. Hence, the view forces one into a kind of error-theory about correct epistemic rules. One consequence of this is having to reject what strikes me as a rather intuitive description of cases of defeat by higher-order evidence: initially it was rational for a subject to apply a given, perfectly correct, epistemic rule, but once the defeating evidence comes it, it is no longer rational for her to do so. For instance, if I acquire strong evidence that I am currently suffering from hypoxia, it is no longer rational for me to believe a given proposition as a result of a sequence of applications of *modus ponens*, even if believing it on such a basis was perfectly rational before I acquired the evidence. But on the Über-rule view, a rule that it is no longer rational to apply couldn't ever have been correct in the first place.

I very much doubt whether there is any finite, informative way of expressing the Über-rule. Recall the kinds of problems faced by attempts to build a clause for higher-order defeaters into familiar-looking rules like *Perception*. *Perception** urged one to follow *Perception* with exceptions: it tells one to follow *Perception* except when one has

[28] Hence, admitting the possibility of an Über-rule comes with the cost of admitting that epistemic rationality falls apart in a range of circumstances. But this may not in itself be objectionable, since anyone who holds on to a principle along the lines of *Higher-order defeat* is pressed to admit that there are such cases. Consider, for instance, evidence that whatever doxastic state one adopts, one is almost certain to commit some cognitive error. It seems that there simply cannot be any rational way of responding to such evidence, for the evidence has defeating force with respect to any attempt to take it into account.

[29] Note that this doesn't guarantee that the following couldn't happen: one can acquire evidence, in some circumstances C_i, that the recommendations made by the rule in other circumstances C_j are incorrect. Hence, one can have evidence that what is in fact a correct Über-rule is incorrect. Someone might worry that this in itself is a problem: how can it be rational to be guided by a rule that it is rational to take to be, overall, incorrect? However, it is at least less clear whether evidence that a rule makes incorrect recommendations in *other* circumstances has any defeating force with respect to the recommendations made by the rule in one's present circumstances.

evidence that the recommendation made by *Perception* is incorrect, or evidence that one is susceptible to making cognitive errors in applying *Perception*. But of course, it is possible to acquire evidence that *Perception** itself is incorrect, or that one has committed an error in attempting to apply *Perception**. Of course, one could invoke a new rule, *Perception***, that recommends following *Perception** with exceptions, but this seems like a project with no endpoint.

One might wonder whether it is even possible for subjects to genuinely follow the Über-rule. Here proponents of the Über-rule view could try to bring simple epistemic rules of the kind we take ourselves to follow to the rescue, for they may act as good heuristics for approximating the Über-rule. The thought is that by being guided by and following rules that are ultimately incorrect, a subject could also be following the Über-rule. After all, there is no reason to think that subjects couldn't follow numerous epistemic rules at once, or that a subject couldn't make mistakes in following an epistemic rule. But even granting this, an Über-rule seems to slot very awkwardly into the theoretical role that epistemic rules are supposed to occupy within the rule-driven picture of epistemic cognition. In effect, the Über-rule view is in danger of compromising the kinds of motivations epistemologists routinely have for adopting the picture in the first place.

To see why, let me digress a bit and ask what makes certain epistemic norms and rules the correct ones, the ones that can be employed to form epistemically rational and justified beliefs. An answer to this question would constitute at least a partial answer to the question of what makes justified beliefs justified. By far the most popular answers have appealed to some sort of connection between justification and the goal or aim of belief, which most take to be truth.[30] Earl Conee, for instance, writes that "the existence of some special intimate connection between epistemic justification and truth seems to be beyond reasonable doubt".[31] A popular way of establishing the connection is to view justification as falling out of practically rational or reasonable means by which to pursue the goal of belief.[32] Hence, the rough idea is that our beliefs aim at truth (or, perhaps, knowledge), and justification is a means to get there. Justified beliefs are good relative to the goal of belief.

Whatever the truth (or knowledge) connection amounts to, rules such as "Believe *p* just in case *p* is true" won't serve as the kind of means for pursuing truth that we are looking for, as such rules don't offer enough guidance as to what exactly it is that one should do in order to believe the truth.[33] What we are looking for is something by means of which true belief and knowledge can be obtained. And it is worth noting that the idea that correct epistemic norms must be capable of offering guidance has been popular independently of appeals to some sort of truth-connection.[34] Now, the problem for the Über-rule view is that an Über-rule just don't seem like the kind of rule that can offer genuine guidance. For one, it cannot even be expressed as a set of finite, informative

[30] See, for instance, Cohen (1984), Conee (1992), Goldman (1979), Kornblith (1993), Wedgwood (2002).
[31] Conee (1992: 657).
[32] See, for instance, Chisholm (1982: 4), Bonjour (1985: 7-8), Conee (1992), David (2001) and Wedgwood (2002).
[33] Cf. Wedgwood (2002: 276).
[34] See, for instance, Pollock & Cruz 1999).

generalisations.[35] It differs from all of the candidate rules that epistemologists have formulated and taken actual subjects to be guided by. Even if one argues that subjects manage to genuinely follow the Über-rule by employing more ordinary kinds of epistemic rules as heuristic guides, the fact remains that they need guidance to follow the Über-rule itself. Hence, the Über-rule is a very awkward candidate for a rule that is itself supposed to play the role of offering genuine guidance.

Resorting to an Über-rule seems to force one into a genuine kind of *epistemic particularism*: there are no useful, informative generalisations about what the epistemically rational ways of managing one's beliefs are.[36] A belief is rational just in case it results from following the Über-rule, but I have argued that we have no reason to expect the rule itself to be expressible in any finite, informative way. At most what we can hope for is rough approximations that track the more ordinary kinds of epistemic rules. Initially one may have at least hoped that facts about epistemic rationality wouldn't be as messy as they come out to be on the Über-rule view.

I don't take myself to have provided a knock-down argument against the Über-rule view. But I do hope to have showed that it involves some very controversial commitments. It is wedded to an error theory about correct rules. It pushes one towards a deep kind of epistemic particularism on which there are no useful, informative generalisations about what the epistemically rational ways of managing one's beliefs are. And, perhaps most important of all, it is not clear whether Über-rules are at all fit to play the kind of guidance-role that proponents of a rule-driven picture commonly require (correct) epistemic rules to play.

---

[35] The phenomenon of higher-order defeat also raises worries about whether any rule that urges one, for instance, to cease inferring by logically valid rules in certain circumstances can guide one towards true belief or knowledge: wouldn't a rule that simply urges one to always infer by valid rules do a much better job? I think there is much to ponder here, but will set the question aside.

[36] Those writing on the subject of moral particularism sometimes take it for granted that reasons for belief behave in a particularist way – in effect, an argument not infrequently given in favour of moral particularism is that particularism is true of reasons in general, including reasons for belief. Moral particularism is usually stated as the view that the moral relevance of a feature or property of an action is not invariant. Often this is split up into a two-part claim: first, there is no invariant way in which features contribute to moral reasons, and second, there is no invariant way in which moral reasons contribute to determining what the morally right thing to do is (see, for instance, Berker 2007). At first sight it may look as though such particularism has got to be true of reasons for belief. Perhaps, for instance, its perceptually seeming to one as if *p* is a (prima facie) reason to believe *p*, but this reason can be defeated by learning, for instance, that one is hallucinating. The issue may rest on exactly what one means by "invariant", but it is far from clear to me that ordinary defeaters push epistemologists towards particularism. Though those who give the notion of a reason a central role in their theories tend to think that the rationality of a doxastic state regarding a proposition *p* is not determined solely by the overall set of reasons that weigh either in favour or against *p*, they tend to think that it *is* determined in an invariant way by the set of reasons together with the set of defeaters (which themselves are reasons, but not ones that need to directly weight in favour or against *p*). Its seeming to one as if *p*, for instance, is always at least a prima facie reason in favour of believing *p*. Indeed, there is a sense in which the views of epistemologists such as John Pollock are paradigm forms of epistemic generalism: given the total set of reasons (and defeaters) present, it is possible to state a general, fairly simple rule for how they combine to determine an overall epistemic verdict about what one should believe (indeed, a large part of Pollock's life project in epistemology was the attempt to formulate such a rule). It is no surprise that Pollock is a proponent of the idea that to be justified, a belief must be licensed by the correct epistemic rules or norms, and he certainly doesn't claim that it is impossible to formulate such norms (see, for instance, Pollock & Cruz 1999).

At this point it is worth asking whether there might be other theories that reject the multiple-tiered theory of justification described in the beginning. Recall that what created the need for Condition 2 – thereby also creating the puzzle discussed at the outset of there being incompatible but correct epistemic rules – was the thought that it is possible to employ a perfectly correct epistemic rule while having evidence that one has failed to do so. In a search for a single-tiered theory of justification, one might suggest giving up the idea that there is a fixed set of correct epistemic rules (a set that might consist solely of an Über-rule). Indeed, one could suggest that what makes a rule correct for a subject at a time is just its being rational for the subject to believe that it has some desirable feature X – perhaps its being rational for her to believe that it is a good means for pursuing truth or knowledge, or even its being rational for her to believe that it is correct. The hope is that in cases of higher-order defeat it is never rational for the subject to believe, of the rule(s) she has employed in forming the relevant belief, that those rules have feature X.

But unfortunately, the proposed account still allows for employing a correct rule while having evidence that one has failed to do so. For instance, the subject's belief may be the output of a rule R that is perfectly correct (because it is rational for her to believe R to have feature X), while she has strong evidence that it is the output of some other, incorrect, rule. Rather than an account that says that a belief is rational or justified just in case it is the result of applying rules that it is rational for the subject to believe to have some property X, perhaps what is needed is an account that says that a belief is rational or justified just in case it is rational for the subject to believe that it is the result of correctly applying epistemic rules with property X.

The resulting theory implies that a belief can be justified even if it is produced by a seemingly terrible rule – indeed, even a rule that it is rational for the subject to regard as terrible. For the subject may rationally but incorrectly take her belief to have come about by the employment of some other rule. And for pretty much any rule, there may be circumstances in which it is rational to believe it to have the desirable property X. So on the proposed account, there are hardly any rules that couldn't produce justified belief. Few epistemologists would be willing to go this far. And of course, the theory faces a rather obvious worry of circularity: rational belief is being characterized in terms of what it is rational for a subject to believe. Even if a fully reductive account or analysis of epistemic rationality or justification cannot be achieved, one might at least hope for a theory that uses a different, even if closely related, notion.

Let me now turn to views that endorse a two-tiered theory of justification. I begin with a view on which defeat by higher-order evidence creates genuine epistemic dilemmas.

V Epistemic binds and dilemmas

One type of reaction to the puzzle outlined above is to concede that defeat by higher-order evidence gives rise to epistemic binds, situations in which it is simply impossible to act in a reasonable or rational manner. Such binds might be created by the incompleteness of the correct epistemic rules, or by their incompatibility. The view I want to discuss now is one on which the correct epistemic rules are incompatible: cases of defeat by higher-order evidence present genuine *epistemic dilemmas*, situations in

which a subject is condemned to doing something she ought not to do and hence, to falling short of epistemic rationality. The view rests on the following two assumptions:

> **1)** If a subject has evidence that her doxastic state is flawed, then the state fails to count as epistemically rational, and she ought to revise it.
>
> **2)** If an epistemic rule is correct, and the rule tells one to believe *p* in circumstances C, and one is in fact in circumstances C, then one ought to believe *p*.

The first of these assumptions more or less follows from a desire to accommodate defeat by higher-order evidence (and from the principle *Higher-order defeat*). It is the second that is distinctive of the epistemic dilemmas view (ED-view). The thought is that the ought attaching to a correct epistemic rule is never defeated in the sense of losing its potency, but the presence of certain types of evidence can create another, competing ought.[37] Failing to revise beliefs in the presence of evidence that they are flawed is a sufficient condition for epistemic irrationality, as there is a correct epistemic norm or rule in place telling one to give up such beliefs. But failing to act in accordance with a correct epistemic rule is likewise a sufficient condition for epistemic irrationality. In situations involving higher-order defeat, two rules that normally happily coexist pull in different directions, neither being overridden by the other. The subject is doomed to epistemic irrationality.

I don't want to rest my case against the ED-view on an argument against the existence of epistemic dilemmas. Rather, what I want to draw attention to is that it is not clear whether the view allows saying the right things about defeat by higher-order evidence after all. Presumably, proponents of defeat by higher-order evidence would want to say that at least in a lot of cases subjects *ought*, all things considered, to revise their beliefs in the presence of higher-order defeaters. Minimally, they would want to maintain that even if a subject's doxastic states are doomed to falling short of full rationality, there is often a best option, an option that takes one closest to rationality or justification.[38] However, I am very skeptical of whether there is a stable ED-view that can make sense of such overall oughts.

Here is a toy picture. Assume that there are degrees of epistemic rationality, and being rational, full-stop, requires being rational to a maximal degree, which means abiding by all correct epistemic rules or norms. There are also degrees of irrationality, but by contrast, the threshold for irrationality is non-maximal (indeed, one might even think that failing to be fully rational entails being irrational). Where a subject's doxastic state falls on the rationality/irrationality scale depends on the extent to which correct epistemic rules are violated. Determining the extent to which correct rules are violated will involve

---

[37] I don't here have in mind a view on which two different *kinds* of oughts compete in cases of defeat by higher-order evidence: one $ought_1$ to be in doxastic state S, but one $ought_2$ not to not be in S. Indeed, the kind of view I want to in the end defend is compatible with different kinds of epistemic oughts pulling in different directions in cases of defeat by higher-order evidence. Instead, I take pure epistemic dilemmas to be ones in which one $ought_1$ to be in doxastic state S, but one $ought_1$ not to not be in S. Thanks to Carrie Jenkins for helping me clarify this.

[38] Christensen (2007a, 2010a) at least comes close to a view on which there are epistemic dilemmas, but even in situations involving such dilemmas, certain doxastic responses are better than others.

some complicated formula that weighs different epistemic rules. The precise details won't matter. What matters is the thought that the less irrational one is the better: all things considered, a subject ought to manage her doxastic states in ways that minimize her irrationality score.

Assume that Rule 1 and Rule 2 are correct epistemic rules. Take a situation in which Rule 1 tells a subject to believe *p*, and Rule 2 tells her to suspend judgment in *p*. The subject ought to believe *p*, but she ought also to suspend belief in *p*, and she cannot do both. The kind of result needed is that abiding by Rule 2 (for instance) constitutes a lesser degree of epistemic irrationality. Perhaps, in the grand scheme of things, Rule 2 bears more weight than Rule 1. But it is reasonable to ask what underlies such facts about which violation would constitute the least degree of epistemic irrationality. It is clear that these questions cannot be answered just by appeal to the correct rules. After all, there is nothing built into Rule 1 or Rule 2 as such that would determine which takes priority in situations of conflict. If Rule 1 had an exception built into it of the sort telling one to do what Rule 2 says in certain situations, we wouldn't be dealing with a genuine epistemic dilemma. And it won't help to introduce an extra rule telling one what the best way of resolving such situations of conflict would be. Assume that Rule 3 tells one to follow Rule 2 in situations of conflict. In so far as one ought to instantiate the least possible degree of epistemic irrationality, the result needed is that one ought to follow Rule 3 – and moreover, that the ought attaching to Rule 3 is somehow weightier than the ought attaching to Rule 1. But the problem is that Rule 3 is just another rule. The overriding ought to abide by Rule 3, thereby violating Rule 1, cannot be created by Rule 3 alone.

To solve this problem, it looks like the correct rules must come with weightings determining the strengths that the oughts attached to them have in situations of conflict. But then the worry arises that we are dealing with an unstable position, having taken on a defining commitment of a view to be discussed below, one that orders epistemic rules in a hierarchical fashion. Indeed, unless the weightings are thought of as determining relations of priority between rules and the oughts attached to them, it is difficult to see what could ground the overall oughts that we are looking for. For let us even concede that the weightings determine degrees of rationality. But why ought a subject to instantiate the least degree of irrationality possible given her situation? One might retort that this objection is based on a misunderstanding. Epistemic rationality just *is* following the correct rules. In addition to the correct epistemic rules, one certainly doesn't need an extra rule telling one to be epistemically rational – what one ought to do is handled by the correct epistemic rules. But the problem is precisely that in situations of conflict, the correct rules *don't* tell one what to do, not unless we are able to extract some overriding, all-things-considered oughts from the way they are weighted or ordered. And besides, it doesn't follow from the fact that one ought to do what the correct rules tell one to do that when these rules come into conflict, one ought to aim for the smallest possible violations. The desirability of instantiating some X to a full degree doesn't generally entail that when achieving X to a full degree is beyond reach, one ought to achieve X to the highest degree possible.

I have expressed scepticism about whether genuine epistemic dilemmas can be reconciled with the idea that even when a subject cannot be fully rational, there may be a most rational option available to her that she ought to choose. To accommodate this idea

proponent of epistemic dilemmas must adopt the kind of hierarchical ordering of epistemic rules that I will now discuss.

VI Hierarchies

The hierarchy view maintains that though cases of defeat by higher-order evidence create situations in which correct epistemic rules make conflicting recommendations, these situations don't constitute epistemic dilemmas, for one of the conflicting rules will overrule or override the others. A particularly welcome result would be that rules pertaining to defeat override the relevant first-order inductive and deductive rules. Take, for instance, a case in which a subject acquires strong but misleading evidence that she has been given a drug that makes her mess up in even the simplest of inferences, while making it seem to her as if she is inferring by perfectly valid rules. The thought is that in such circumstances a perfectly correct, valid inference rule may be overridden by a rule telling one to give up belief in the presence of sufficiently strong evidence that the cognitive process producing the belief is flawed.

To get such overruling to happen, in addition to a set of correct epistemic rules, we need some sort of relation ordering those rules, a relation determining which rules take priority in situations of conflict.[39] An epistemic system will be comprised of two elements: a set of correct epistemic rules, and an ordering relation on these rules. The former can be thought of as functions from (epistemic) circumstances to doxastic states; the latter can be thought of as a function from (epistemic) circumstances to functions from epistemic circumstances to doxastic states – that is, a function from circumstances to epistemic rules. As such, the ordering relation is a kind of meta-rule, a rule telling one which rule to follow.

At this point one may wonder what the difference between the hierarchy view and the Über-rule view amounts to. After all, given any hierarchy of epistemic rules, one can construct a super-rule based on the hierarchy as follows: for any possible situation, let the super-rule recommend doing just what the hierarchy as a whole would recommend. Given that the epistemic system comprising the hierarchy and the super-rule would fully agree about which doxastic responses are rational in any possible circumstances, isn't the hierarchy view equivalent to the Über-rule view? While it may be true that the two views agree on what doxastic responses would be rational in any possible situation, the structures generating these verdicts are very different. And there are various issues the views disagree about. In particular, the hierarchy view is able to recognize a distinction between which rules it is rational to employ on a given occasion and which rules are correct: a rule R can be perfectly correct and rational to follow even if there are situations in which it is not rational to follow R, since in those situations R is dominated by a higher-level rule. But recall that on the Über-rule view it can never cease to be rational to follow a correct rule. The hierarchy view posits a plethora of correct rules and a relation ordering them (relative to circumstances), the Über-rule view posits exactly one.

[39] Though I speak of hierarchies of epistemic rules, thinking in terms of an image of a fixed hierarchy of rules may not be accurate. I want to allow for a rule R to override another rule R' in certain circumstances, but for R' to override R in others.

Can the hierarchy view do with a single meta-rule that orders the first-order epistemic rules? Here the hierarchy theorist faces an important decision point. Recall that the need for a meta-rule ordering the correct first-order rules arose because it was assumed that correct epistemic rules could continue to make recommendations in situations in which one has higher-order evidence with defeating force with respect to those recommendations. But to prevent having to take refuge in a meta-meta rule ordering candidate meta-rules, nothing analogous had better be the case for the meta-rule itself. For imagine that the meta-rule could recommend following rule R_1 (rather than R_2) in a situation in which the two rules conflict, but in which one has evidence that any ordering giving priority to R_1 over R_2 is flawed. This would look very much like a higher-order defeater for the meta-rule itself. Or, take a case in which instead of having evidence that the ordering delivered by the meta-rule is incorrect, one has evidence that one is highly likely to incorrectly apply the meta-rule, giving priority to the wrong epistemic rule.

To prevent such situations from arising, one could stipulate that the meta-rule is special: unlike first-order epistemic rules, it will never make a recommendation in the presence of evidence with defeating force with respect to that recommendation. But by now the thought that a rule is thus immune to defeat should sound familiar, for recall that it is a distinctive feature of Über-rules! If the meta-rule is invincible when it comes to defeat, then in effect, it is an Über meta-rule. Hence, on the kind of hierarchy view under consideration, the need for an Über-rule is merely pushed up one level. Not only does the resulting view buy into core commitments of the Über-rule view, thereby being susceptible to similar problems, but it also looks like an awkward compromise between two different approaches.

The other option is to admit that higher-order defeat can happen at the level of meta-rules: it is possible for circumstances to arise, for instance, in which a meta-rule M_1 recommends following some first-level rule R_1 rather than R_2, but there is evidence that any such recommendation is incorrect. Perhaps this evidence points to the rationality of following some further rule R_3 instead. If such cases are possible, then it looks like the hierarchy theorist will have to resort to meta-meta rules, for we now face a situation at the level of meta-rules that is similar to the one we faced at the level of first-order epistemic rules: a meta-rule recommends following a first-order rule R_1, but acting in accordance with the recommendation is irrational because of a higher-order defeater. In so far as one now ought not to adopt the doxastic state recommended by R_1, one ought to follow some other rule instead. Then, there must be an alternative meta-rule M_2 making a recommendation that conflicts with the recommendation made by M_1. But if there are two meta-rules M_1 and M_2 making conflicting recommendations, for reasons outlined above we need a meta-meta rule to tell us which one to follow. Just as meta-rules can be represented by functions from circumstances to first-order rules, meta-meta-rules can be represented by functions from circumstances to meta-rules.

Allowing for multiple correct meta-rules that make conflicting recommendations in certain situations looks like the beginning of an infinite regress. Hence, the only alternative to admitting that the meta-hierarchy terminates in an Über-rule at some level seems to be admitting that it doesn't terminate at all. In an infinite meta-hierarchy, conflicts between epistemic rules can arise at any level. I for one am far from comfortable with the thought that the correct epistemic system consists of such an infinite hierarchy of

epistemic rules, though I admit to not being able to offer decisive argument against such hierarchies. Here are a couple of qualms. First, there is no upper bound to how far up the meta-hierarchy one would have to climb in some situations before finding a recommendation made by a meta-rule that is not defeated. As a result, it is doubtful whether finite beings could be guided or governed by such complex epistemic systems. Again, can the infinite hierarchy as a whole play the guidance-role that correct epistemic rules were initially supposed to play? Moreover, it is not clear how different in the end the hierarchy view is from the kind of epistemic particularism sketched in connection with the Über-rule view. Because of the multiple forms that higher-order evidence can take, it is far from clear to me whether the infinite hierarchy would exhibit patterns that could be stated in finite generalizations. As a result, it is not clear if we can extract from the hierarchy useful, informative generalizations about epistemic rationality – at least not ones that could be stated in a finite language. If systematic higher-order defeat of the kind that respects principles like *Higher-order defeat* formulated above is a real epistemic phenomenon, facts about it look to be very messy.

The proponent of the hierarchy view faces a choice: either make one of the meta-rules into an Über-rule, a rule such that its recommendations can never be defeated, or allow for an infinite hierarchy of meta-rules, with numerous correct rules at each level that can come into conflict with each other. Neither option strikes me as a triumphant way out of the puzzle created by trying to accommodate higher-order defeat within the context of a two-tiered theory of justification.

VII The limits of defeat

I have looked at possible ways of responding to the puzzle sketched above. Of the views that endorse systematic defeat by higher-order evidence, the strongest contenders appear to be the Über-rule view and a view that endorses an infinite hierarchy of correct epistemic rules. A theme common to both approaches is a kind of epistemic particularism, which may not be reconcilable with common commitments of the rule-driven picture of epistemic cognition. Hopefully even those who don't share my qualms about these views will find the exposition of their commitments useful. But assume that I am right that there is no elegant way of accommodating systematic defeat by higher-order evidence. I want to end on a more speculative note, outlining what I take to be the correct response to the initial puzzle.

Ultimately, I think the right conclusion to draw is that the desire to accommodate limitless defeat lands one into a paradoxical predicament with no happy resolution. The predicament is created by the following two assumptions. First, for any epistemic rule or principle, there are possible situations in which one acquires evidence that a doxastic state that is in fact an output of that rule is flawed. And second, in such circumstances epistemic rationality calls for revising the doxastic states in question: there is some correct epistemic rule or principle dictating that it would be irrational not to do so. The conclusion I want to draw is that there is no non-paradoxical notion of justification or epistemic rationality that can accommodate these ideas.

I don't see any grounds for questioning the first assumption. Nor do I think that one should jettison the very idea of epistemic rationality as inherently incoherent. Instead, I think the second assumption is the culprit. It may come as a surprise that in some cases

a state can be perfectly epistemically rational even if one has what would seem like strong evidence for thinking that it is not. In particular, in so far as there is such a thing as a correct inductive policy or epistemic system, it can be rational to follow the recommendations of that policy or system even if one possesses evidence that in doing so one has committed a rational error. Such a conclusion doesn't rest on anything like an externalist epistemology. It rests merely on a desire to avoid paradox. Hence, I take the considerations put forth above to constitute at least a tentative, non question-begging argument for a kind of externalism.

So, for instance, it may be rational to conditionalise a correct prior credence function on what would look like evidence for thinking that that credence function is incorrect – or even evidence that conditionalisation is not the right rule for taking new evidence into account. Sometimes this may result in a situation in which it is rational to assign a high credence to one's epistemic rules or policies being incorrect. Now, someone might be inspired by remarks made by David Lewis to reject such a position as incoherent.[40] For how could a correct epistemic system recommend one doxastic response while recommending the belief that that doxastic response is incorrect? Wouldn't it be recommending that one φ while recommending that one adopt in its place a system that doesn't recommend that one φ, thereby issuing incompatible recommendations?[41] I think the correct response to this is that it would do no such thing. Recommending that one believe that a rule is flawed is not tantamount to recommending that one stop following the rule. That one should believe that one shouldn't φ doesn't entail that one shouldn't φ.[42]

Thinking about defeat easily leads to over-inflating the notion of epistemic rationality. There is an intuition that subjects ought to revise their doxastic states in the presence of evidence that those states are flawed, and that they are criticisable for not doing so. I think these intuitions correct. Moreover, I think that subjects who fail to revise their beliefs in putative cases of defeat are criticisable from an epistemic point of view: they are being *unreasonable* by failing to take into account evidence about their own cognitive imperfections, thereby manifesting dispositions that are bad from the perspective of acquiring knowledge or true belief.[43] But I very much doubt whether there is a non-paradoxical notion of epistemic rationality that marches step in step with such criticisability. There are epistemic oughts that a subject can violate without thereby being epistemically irrational, or failing to meet the criteria for justification.[44]

---

[40] See Lewis (1971).

[41] Cf. Schechter (2011).

[42] I suspect that some such confusion is, for instance, driving people to argue against the equal weight view of disagreement on the grounds that in some circumstances the view would urge one to give up belief in its own correctness.

[43] For more on my notion of reasonableness, see Lasonen-Aarnio (2010).

[44] Many thanks to Ville Aarnio, Sara Aronowitz, Dave Baker, David Christensen, Daniel Drucker, Adam Elga, Hartry Field, Alan Gibbard, John Greco, John Hawthorne, Carrie Jenkins, Marja-Liisa Kakkuri-Knuuttila, Markus Lammenranta, David Manley, Sarah Moss, Anders Nes, Peter Railton, Josh Schechter, Eric Swanson, Jonathan Vogel, Tim Williamson, and numerous members of the audience at the 2011 Mark Shapiro Philosophy Conference at Brown University, at a research seminar at the University of Oslo, and at an Eastern APA session in December 2011.

Bibliography

Berker, Selim (2007). Particular Reasons. *Ethics* 118(1): 109-139.

Boghossian, Paul (2008). Epistemic Rules. *Journal of Philosophy* CV (9): 472-500.

Bonjour, Lawrence (1985). *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Chisholm, Roderick M. (1982). *The Foundations of Knowing*. Minneapolis: University of Minnesota Press.

Christensen, David. (2010a). Higher-Order Evidence. *Philosophy and Phenomenological Research* 81(1): 185-215.

---------- (2010b) Rational Reflection. *Philosophical Perspectives* 24(1): 121-140.

---------- (2009). "Disagreement as Evidence: The Epistemology of Controversy," *Philosophy Compass* 4 (5):1-12.

---------- (2007a). Epistemic Self-Respect. *Proceedings of the Aristotelian Society* 107: 319 - 337.

---------- (2007b). Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals. *Oxford Studies in Epistemology* 2: 3 - 31.

---------- (2007c). Epistemology of Disagreement: the Good News. *Philosophical Review* 116 (2): 187 - 217.

Cohen, Stewart (1984). Justification and Truth. *Philosophical Studies* 46 (3): 279-295.

Conee, Earl (1992). The Truth Connection. *Philosophy and Phenomenological Research* 52 (3): 657-669.

David, Marian (2001). Truth as the Epistemic Goal. in Steup, M. (ed) *Knowledge, Truth and Duty: Essays on Epistemic Justification, Responsibility and Virtue*. Oxford: Oxford University Press, 151-169.

Elga, Adam (Unpublished). Lucky to be Rational.

Field, Hartry (2000). Apriority as an Evaluative Notion. in P. Boghossian and C. Peacocke, eds., *New Essays on the A Priori*. New York: Oxford University Press, 117-49.

Goldman, Alvin (1979). "What Is Justified Belief?" in G. Pappas (ed.), *Justification and Knowledge*, Dordrecht: Reidel, 1-23. Reprinted in A. Goldman, *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, Cambridge, MA: MIT Press (1992).

--------- (1986). *Epistemology and Cognition*. Cambridge, MA and London, England: Harvard University Press.

Hawthorne, John (2004). *Knowledge and Lotteries*. Oxford: Oxford University Press.

Kelly, Thomas (2010). Peer Disagreement and Higher-Order Evidence. in R. Feldman and T. A. Warfield. (eds.), *Disagreement*. Oxford: Oxford University Press, 111-174.

Kornblith, Hilary (1993). Epistemic Normativity. *Synthese* 94 (3): 357-376.

Lasonen-Aarnio, Maria (2010). Unreasonable Knowledge. *Philosophical Perspectives* 24(1): 1-21.

Lewis, David (1971). Immodest Inductive Methods. *Philosophy of Science* 38(1): 54–63.

Pollock, John L. and Joseph Cruz (1999). *Contemporary Theories of Knowledge* (2nd edition). Lanham, MD: Rowman & Littlefield.

Schechter, Joshua (2011). Rational Self-Doubt and the Failure of Closure. *Philosophical Studies* 163(2):429-452.

Sorensen, Roy A. (1988). Dogmatism, Junk Knowledge, and Conditionals. *The Philosophical Quarterly* 38(153): 433-454.

Wedgwood, Ralph (2002). The Aim of Belief. *Philosophical Perspectives* 16, Language and Mind: 267-297.