Epistemic Akrasia

SOPHIE HOROWITZ
Massachusetts Institute of Technology

Many views rely on the idea that it can never be rational to have high confidence in something like, "P, but my evidence doesn't support P." Call this idea the "Non-Akrasia Constraint". Just as an akratic agent acts in a way she believes she ought not act, an *epistemically* akratic agent believes something that she believes is unsupported by her evidence. The Non-Akrasia Constraint says that ideally rational agents will never be epistemically akratic.[1]

The importance of the Non-Akrasia Constraint is brought out by recent literature on "higher-order evidence"—evidence about what evidence one has, or what one's evidence supports. Much of this debate has focused on the question of what to believe when one becomes rationally highly confident that P on the basis of some evidence, E, and then receives further (misleading) evidence to the effect that E does not support P. Although there is disagreement as to whether this new evidence should cause one to reduce confidence in P, the major positions that have been defended thus far in the higher-order evidence debate agree that ideally rational agents should respect the Non-Akrasia Constraint.

But in a number of recent papers, the Non-Akrasia Constraint has been called into question. Some have argued that in cases of misleading higher-order evidence, epistemic akrasia can in fact be rationally required. If you have excellent evidence for P, and excellent evidence that your evidence doesn't support P, the thought goes, perhaps the rational response is to have high confidence in both of these things. The tenability of these pro-akrasia (or "Level-Splitting") views will depend, in part, on whether there is a strong case for the Non-Akrasia Constraint. So what can be said in favor of the constraint?

The goal of this paper is to defend the Non-Akrasia Constraint . . . for the most part. In the first part of the paper I defend the constraint indirectly, by showing that Level-Splitting views face serious challenges: they have counterintuitive consequences regarding the rational response to higher-order evidence, and it is unclear how these views could deliver plausible results in nearby cases. The upshot of this discussion, and of the bulk of the paper, is that we should accept the Non-Akrasia Constraint.

In the last section, I consider a different kind of challenge to Non-Akrasia, based on Williamson's case of the unmarked clock. Looking at this less-paradigmatic type of case gives us reason to think that Non-Akrasia may not be a universal rational requirement. But, surprisingly, I suggest that allowing rational epistemic akrasia

under certain special conditions does not undermine anti-akratic verdicts (or the arguments against Level-Splitting) in paradigm cases.

1. A Paradigm Case

Consider the following story:

> **Sleepy Detective:** Sam is a police detective, working to identify a jewel thief. He knows he has good evidence—out of the many suspects, it will strongly support one of them. Late one night, after hours of cracking codes and scrutinizing photographs and letters, he finally comes to the conclusion that the thief was Lucy. Sam is quite confident that his evidence points to Lucy's guilt, and he is quite confident that Lucy committed the crime. In fact, he has accommodated his evidence correctly, and his beliefs are justified. He calls his partner, Alex. "I've gone through all the evidence," Sam says, "and it all points to one person! I've found the thief!" But Alex is unimpressed. She replies: "I can tell you've been up all night working on this. Nine times out of the last ten, your late-night reasoning has been quite sloppy. You're always very confident that you've found the culprit, but you're almost always wrong about what the evidence supports. So your evidence probably doesn't support Lucy in this case." Though Sam hadn't attended to his track record before, he rationally trusts Alex and believes that she is right—that he is usually wrong about what the evidence supports on occasions similar to this one.

What should Sam believe about who the jewel thief is? And what should he believe about what his evidence supports?

In Sleepy Detective, Sam's epistemic position can be described in the following way: first, he receives some "first-order" evidence, E, which includes the letters and photographs that he was looking through as he worked late at night. E supports a proposition, P: that Lucy is the thief. So after considering E, it seems that Sam should believe P. Then, when Sam talks to Alex, he receives some higher-order evidence, HOE: her testimony about his track record data, which suggests that in situations like this one, he almost always misinterprets his evidence. HOE therefore suggests that the answer Sam came up with is unlikely to be supported by his evidence; since Sam came up with P, HOE suggests that P is unlikely to be supported by Sam's evidence. (Note that in order for this to be true, we don't have to suppose that Sam is *in fact* unreliable; all we need is that HOE supports the proposition that he is.) So HOE supports this proposition: that E does not support P.

What does Sam's *total* evidence support?

Two answers to this question have been most commonly defended in the literature; I will mention them only briefly, and then largely set them aside for the remainder of the paper. According to "Steadfast" views, Sam should maintain his high confidence in P, and also maintain his high confidence that E supports P.[2] The advantage of Steadfast views is that they acknowledge the evidential force of E; the disadvantage is that they seem to ignore HOE. According to "Conciliatory" views, Sam should become much less confident in P, and also become less confident that E supports P.[3] The advantage of Conciliatory views is that they acknowledge the

evidential force of HOE; the disadvantage is that they seem to ignore E. Although Steadfast and Conciliatory views disagree about what Sam's total evidence supports, they share a common assumption: that Sam's confidence in the first-order proposition, P, should line up with his confidence in the proposition that E supports P. Thus, both Steadfast and Conciliatory views respect the Non-Akrasia Constraint in cases like Sleepy Detective.

The view that I will discuss here, which has become more prominent in recent literature, offers a third answer. Call this third view "Level-Splitting". Level-Splitting says that Sam should maintain his high confidence in P, but he should become much less confident that E supports P. In short, he should become highly confident in "Lucy is the thief, but my evidence doesn't support that." Level-Splitting might seem like a happy compromise between Steadfast and Conciliatory views. In accommodating both first-order and higher-order evidence, it has the apparent advantages of both views, and the apparent disadvantages of neither. But in one important respect, Level-Splitting is not a compromise between Steadfast and Conciliatory views. While Steadfast and Conciliatory views both respect the Non-Akrasia Constraint, Level-Splitting denies it.

In the next section, I will turn to discussing some Level-Splitting views that have been defended in recent literature.

2. Level-Splitting Views

The view I'm calling "Level-Splitting" will be picked out by its verdict in cases like Sleepy Detective: that it can be rational, in that type of case, to have high confidence in both "P" and "my evidence doesn't support P." Of course, there are many possible ways to motivate this verdict, and the four authors I consider below arrive at the view in different ways. Looking at these views will shed light on what kinds of epistemological positions are compatible with the Non-Akrasia Constraint.

2.1 Williamson
In his "Improbable Knowing" and "Very Improbable Knowing", Timothy Williamson argues that there can be cases where one knows that P, but it is also highly improbable on one's evidence that one knows that P. In certain cases, Williamson's view is committed to a denial of the Non-Akrasia Constraint: it requires a rational agent to have high confidence in "P, but my evidence doesn't support P."

One example of this type is what Williamson calls the "Long Deduction". Suppose a rational agent comes to know a series of claims and competently deduces their conjunction, C. On Williamson's view, she can come to know C by these means. But suppose further that the rational agent also knows that, since most people's memories and logical abilities are limited, people in her situation often make inferential errors while completing long deductions. It can then be highly probable on her evidence that *she* has made such an error, and thus that she does not know the conjunction. Nevertheless, since she has in fact competently deduced

C, Williamson holds that she does know C. In a case like this, it will be certain on her evidence that C is true—for Williamson, knowledge of C requires assigning C evidential probability 1—but also highly probable on her evidence that she does not know C.[4] So she should be highly confident in C, and highly confident that she does not know that C.

High confidence in "C" and "I don't know that C" is strange, but it's not yet a violation of the Non-Akrasia Constraint. ("I don't know that C" is compatible with "C is highly probable on my evidence", and high confidence in "C, and C is highly probable on my evidence" does not have quite as much of a Moore-paradoxical flavor. I'll say more about less-extreme kind of mismatch in section 2.5.) On a plausible filling out of Williamson's story, however, an ideally rational agent can also be highly confident of both "C" and "my evidence doesn't support C".

We have seen that our rational agent can be rationally highly confident that she does not know C because of doubts about inferential mistakes; so she must be rationally highly confident that she has made some such mistake. What kind of mistake should she think she has made? Usually, when we make mistakes in completing long deductions, we reach false conclusions. But our rational agent can't think that this is what's happened; after all, she knows that the conclusion she reached is C, and she is certain that C is true. So she must think she has made some other kind of inferential error: one that would result in a true belief, but not knowledge.

Whether a rational agent will become highly confident of both "C" and "my evidence doesn't support C", in these circumstances, will depend on what *kind* of inferential error she thinks she has made.[5] Some kinds of error are compatible with C's being supported by her evidence. For example, if it's a mathematical deduction, she might think she made several calculation errors that canceled one another out, luckily resulting in a true belief. Or she might worry that, even though her evidence (whatever it is) supports C, there's a good chance that her belief isn't properly "based on" that evidence. In these cases, the agent should believe C is true, and supported by her evidence—but nevertheless she might be rationally highly confident that her belief in C does not amount to knowledge.

Errors of a second type would make it likely that C is *not* supported by her evidence. For example, she might think she has misidentified her evidence, and that the evidence she could have mistaken it for does not support C. Or she might think she has made a mistake in identifying what her evidence supports: although her evidence doesn't support C, C is true, and she (luckily, but irrationally) believes it. If an agent has reason to believe that she has made an error of this second type, it is plausible that she should become highly confident that her evidence doesn't support C.

Williamson agrees that there can be cases in which an agent is unsure of what her evidence is, or of what her evidence supports. And it is possible that sometimes one can be not only unsure, but significantly mistaken: as Williamson writes in his "Very Improbable Knowing", "one's evidence can be radically misleading about one's own present epistemic position."[6] In other cases, Williamson allows that the evidential probability of P can be quite high, while it is also highly probable that

the evidential probability of P is quite low.[7] So it's plausible that, on his view, there can be cases where an agent is rationally highly confident that C (as a result of competent deduction) but also rationally highly confident that her evidence doesn't support C (because she is rationally highly confident that she has made an error of the second type above).

One such case might be Sleepy Detective: while it's highly likely on Sam's first-order evidence that Lucy is the thief, it's also highly likely on Sam's higher-order evidence that that he has made some kind of error in evaluating his evidence. If Sam has reason to believe that his error is of the more problematic type—one that would lead him to believe something *unsupported* by his evidence—it seems that on Williamson's view, Sam should be highly confident of both "Lucy is the thief" and "my evidence doesn't support that."

2.2 Lasonen-Aarnio

Maria Lasonen-Aarnio, in her "Higher-Order Evidence and the Limits of Defeat", also argues that one can have rational high confidence in both "P" and "my evidence doesn't support P" in cases of misleading higher-order evidence. Lasonen-Aarnio's reasons for thinking so arise from a puzzle that emerges when one tries to artic-ulate precisely *how* higher-order evidence could defeat first-order attitudes. Prima facie, she points out, it appears that higher-order defeat requires rationality to give "conflicting recommendations". In Sleepy Detective, for example, one correct epis-temic rule apparently tells Sam to be highly confident in P (since arguably, that's what it is for Sam's first-order evidence to strongly support P). And if higher-order defeat is a genuine phenomenon, there is another correct epistemic rule telling Sam to reduce confidence in P. So Sam is subject to contradictory norms. How could this be?

Lasonen-Aarnio surveys several possible responses to the puzzle, and argues that none is satisfactory. Ultimately, Lasonen-Aarnio suggests that there can be no single coherent, non-paradoxical notion of rationality that accommodates higher-order defeat.[8]

What conclusion should we draw? Lasonen-Aarnio argues that we should give up on the idea of higher-order defeat: sometimes a rational agent can accommodate higher-order evidence in her attitudes about what her evidence supports without revising her confidence in her first-order attitudes. Lasonen-Aarnio writes that "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;" and "[t]hat one should believe that one shouldn't Φ doesn't entail that one shouldn't Φ."[9] So in rejecting higher-order defeat in the way she does, Lasonen-Aarnio also rejects the Non-Akrasia Constraint.

2.3 Wedgwood

A third example of a recent view along the lines of Level-Splitting comes from Ralph Wedgwood's "Justified Inference". On Wedgwood's view, an inference is justified if

and only if, in drawing that inference, one actually manifests a "necessarily rational disposition". He writes,

> [I]t is the real nature of this internal process—the fact that it is a process of competent inference—and not the higher-order beliefs that the thinker has, or even the beliefs that the thinker is justified in having, about the nature of that process, that is crucial to the rationality of the mental event that results from that process.[10]

Beliefs formed through justified inferences will also be justified. And if higher-order evidence doesn't change the "real nature" of one's reasoning process, it also does not rationally mandate that one revise one's confidence in the first-order beliefs formed using that process.

In discussing higher-order evidence, Wedgwood writes that on his view, "you might be in the weird situation of rationally drawing an inference even though it is *not* rational for you to *believe* the inference to be rational."[11] This is true even when you have reason to believe that the inference was *not* rational: Wedgwood writes, "a perfectly rational thinker would continue to draw the inference even if she had (misleading) evidence that she was reasoning incompetently, and even if she entertained serious doubts about whether or not she really was perfectly rational."[12]

So in Sleepy Detective, if Sam's inference to the conclusion that Lucy is the thief was *in fact* an instance of the right kind of mental process, nothing Alex could tell Sam about his own rationality could defeat the justification of Sam's first-order belief. Wedgwood's view thus denies the Non-Akrasia Constraint.

2.4 Weatherson and Coates

Finally, Brian Weatherson and Allen Coates each directly address the question of when it could be rational to believe a conjunction of the form, "P, but my evidence doesn't support P." Weatherson writes that "[t]he circumstances that could make it reasonable are ones where the agent has good evidence that [P], but also has misleading evidence about the quality of her evidence. ... By hypothesis, each conjunct is well-supported by the evidence."[13] Coates writes, in a similar vein: "Like all evidence, second-order evidence can sometimes be misleading. When it is, we can rationally judge that our belief is irrational even though it is in fact rational, and so we can rationally be akratic."[14] These authors' general thought is that whenever higher-order evidence is misleading, our first-order attitudes should come apart from our attitudes about what we should believe. So both authors take cases like Sleepy Detective to be counterexamples to the Non-Akrasia Constraint.

To defend this sort of verdict, both authors discuss analogies with moral cases. Weatherson, for example, asks us to imagine the student of a Kantian professor who finds herself facing a murderer at the door, wondering whether to lie about the whereabouts of her roommate. Since her professor's arguments were so sophisticated and persuasive, Weatherson argues, the student should *believe* that she shouldn't lie to the murderer; that's what her evidence supports. Nevertheless, lying to the murderer *is* what she should *do*. Weatherson then suggests that the same kind of situation can occur in epistemology: "sometimes what we should believe is different from what we should believe that we should believe."[15]

The purely epistemic cases that are most analogous to Weatherson's moral case are those like Sleepy Detective: situations in which an agent responds correctly to some first-order evidence, and then receives evidence that she has not responded correctly to her first-order evidence. In the course of his paper, Weatherson considers a number of cases that fit this pattern. He writes,

> [T]here are facts about which hypotheses are supported by which pieces of evidence, and ... rational agents do well when they respond to these epistemic facts. ... [T]hese facts retain their normative significance even if the agent has reason to believe that she's made a mistake in following them. That is, if an agent's judgment conforms to the correct norms of judgment, then even if she has evidence that she is not good at judging, she should stick to her judgment.

Coates considers similar cases, and concludes that in such situations "one's total evidence can support both the belief that *p* and the judgment that the belief that *p* is irrational."[16] So on these authors' views, agents in cases like Sleepy Detective are rationally required to be epistemically akratic.[17]

2.5 Extreme Level-Splitting

Williamson, Lasonen-Aarnio, Wedgwood, Weatherson and Coates are thus all committed to the following claim: in cases like Sleepy Detective, we can be rationally required to be epistemically akratic. It is clear that these authors' views will support the verdict that one can be rationally highly confident in P, and also rationally required to be highly confident that one's evidence doesn't support *as high a level of confidence in P as one actually has*. And it is plausible that these authors' views will also support the verdict that in more extreme Sleepy Detective-style cases, one can be rationally highly confident in P *and* rationally highly confident that one's evidence supports very low confidence in P.

For the remainder of the paper, I will focus on the possibility that extreme kinds of mismatch can be rational in paradigm cases of misleading higher-order evidence. The question, then, is whether this extreme view can be right: whether in cases like Sleepy Detective, it can be rational to have high confidence in both "P" and "my evidence doesn't support P", "my evidence supports low confidence in P", or "my evidence supports ∼P". One might hold that these more extreme states are always irrational, but maintain that some more moderate mismatches can be rational. Allan Hazlett, for instance, defends a view on which one can rationally believe P while suspending judgment about whether one's belief in P is reasonable—but one cannot rationally believe P while *believing* that one's belief in P is *unreasonable*.[18] As Hazlett points out, constraints barring the more extreme form do not entail that the moderate form is also unacceptable.

But while it may be possible to accept a more moderate Level-Splitting view, many of the considerations surveyed above clearly support an extreme version. Weatherson's moral analogy, for example, relies on the idea that one can be morally required to take some action (i.e. lie to the murderer) while being rationally required to believe that one should take a different, incompatible action (i.e. tell the truth). The most direct epistemic analogy would be a case in which one believes P, but also

believes that one should believe ~P: a paradigm case, that is, of extreme epistemic akrasia. On Wedgwood's view, the rationality of an inference is entirely a matter of the particular mental process. If that process can occur even when an agent harbors "serious doubts" about its reliability, a natural consequence is that an agent's belief in P can be rational despite her believing that she should not believe P. If these motivations support any kind of Level-Splitting view, they support one on which extreme epistemic akrasia can be rational. And this consequence is clearly a welcome one, to at least some authors; as we've seen, some of the authors quoted above explicitly extend their views to the extreme cases.

If extreme divergence is *irrational*, we can rule out extreme Level-Splitting views completely. Moreover, looking at this strong kind of divergence is worthwhile even for those looking to support weaker Level-Splitting views: seeing why extreme epistemic akrasia is irrational can shed light on whether a moderate Level-Splitting view could be supported, and if so, what kind of desiderata such a view must meet. For present purposes, then, I will focus on Level-Splitting views on which the full range of epistemically akratic attitudes can be rational in cases like Sleepy Detective.

3. Immediate Problems

If Sam follows Level-Splitting's advice, he will come to have high confidence in something like, "Lucy is the thief, but my evidence doesn't support that." This, by itself, already looks strange; the conjunction of the two statements has a somewhat Moore-paradoxical quality.[19] Although statements of this form are not contradictory—both conjuncts will be true whenever one's evidence is misleading— there is something odd about believing or asserting them simultaneously.

Since Level-Splitters claim that Sam is rationally required to have high confidence in both statements, they won't be moved very far by contentions of the conjunction's intuitive oddness. But it's worth pointing out that the oddness extends beyond the combination of attitudes itself: the counterintuitive consequences Level-Splitters must accept are more far-reaching than one might have thought. I will go through a few immediate examples here.

First, suppose that Sam, having reacted as Level-Splitting recommends, reflects on his situation. He is highly confident that Lucy is the thief. And he has some first-order evidence, E, which he is highly confident does *not* support the proposition that Lucy is the thief. What should Sam think about his epistemic state? If Sam takes both "Lucy is the thief" and "my evidence doesn't support Lucy" as premises, it seems that he can engage in some patently bad reasoning.[20]

More specifically, first: what should Sam think about his first-order belief that P? Upon reflection, Sam might naturally wonder how he came to have this particular true belief. Usually, we come to have true beliefs by correctly evaluating our evidence. But Sam doesn't think that this is what's happened in his case: he believes that his evidence *doesn't* support P. So perhaps he should just think that he got lucky. "Despite the odds," he should think, "I've come to the truth about P!"

The oddness of Sam's attitude toward his belief that P might look even stranger if we imagine another situation in which Sam *first* receives some higher-order evidence

to the effect that he is going to be unreliable in assessing some evidence, E, in the future. Sam should then become highly confident that whatever judgment he makes at that future time will be false; after all, that's usually what happens when we make irrational judgments. But when the future time comes and Sam assesses E, finding himself inclined to judge that P on its basis, what should he think? According to the Level-Splitting view, he should be highly confident that P. "I thought I was going to judge falsely," Sam might say to himself, "but I must have I lucked out! I judged that P, and P is true."

Second, what should Sam think about his first-order evidence, E? Well, he is rationally highly confident that E bears on P; E is all of his evidence about the jewel thief case, and he knows that his evidence will strongly support the guilt of one of the suspects. This means that it is rational, given E, to have high confidence in the guilt of one suspect and low confidence in each of the others: for each suspect, Sam's rational confidence that she is guilty, on E, should either be quite high or quite low. On the basis of his higher-order evidence, Sam is rationally highly confident that E *doesn't* support P—that Lucy is the thief. So, Sam should think, most likely he should have quite low confidence in P. But because of his first-order evidence, Sam's confidence in P is quite *high*. It seems that he can then put the pieces together and reason in the following way: "P is true. But all of my evidence relevant to P does not support it. It supports *low* confidence in a true proposition, P, and therefore high confidence in a false proposition, ~P. So, E is misleading."

Is this a legitimate way to draw the conclusion that one's evidence is misleading? To better see what's wrong with Sam's reasoning, consider an analogous case: Sam is driving in to work one morning, and plans to evaluate some evidence, E, when he gets there. He hasn't seen E yet, but as he is driving, he rationally thinks to himself: "Whatever E is, it is most likely not misleading—after all, good evidence usually supports the truth." When he walks in the door, Alex tells him (falsely) that she has spiked his coffee with a reason-distorting serum. "That's too bad," Sam thinks. "Now I won't be able to evaluate my evidence correctly; I'm unlikely to find out the truth about the jewel thief." Then Sam opens the door to the evidence room, and sees E before him. E, in fact, supports P. Just as in Sleepy Detective, Sam evaluates E rationally, concludes that P, and takes his higher-order evidence to indicate that E does not support P. "Aha!" he exclaims, "P is true! But this evidence doesn't support P; it's misleading. That's too bad." Then Sam thinks another minute, and smiles. "Actually, it's a good thing Alex spiked my coffee. If I had evaluated my evidence correctly, I would have ended up with a false belief."

Sam's reasoning here does not look rational: he should not be able to conclude that E is misleading in this way. But we should be careful before condemning his argument: there are circumstances in which it *is* rational to conclude that you have misleading evidence. For example, it can be rational to believe that a particular subset of your evidence is misleading with respect to a particular proposition. This might happen when you receive new, more conclusive evidence, showing that your original evidence pointed in the wrong direction. So you can rationally believe P, but also rationally believe that part of your evidence supports ~P. In this type of situation, we can tell a sensible story about why your belief state is stable. By basing

your belief on your *total* evidence, you have avoided being misled. So although you realize that you have misleading evidence, there is no pressure to revise your beliefs.

It can even be rational, in some cases, to conclude that your *total* evidence is misleading.[21] For example, in a lottery case, you can be rationally highly confident that one of the propositions of the form "Ticket n will lose" is false, though you are also rationally highly confident in each of these propositions. Here, we can similarly tell a story about why your belief state is stable. Since you cannot identify which ticket proposition is false, you cannot identify the particular way in which your evidence is misleading. There is no particular other belief state that you think would be more accurate than your own belief state; therefore, there is no way for you to avoid being misled. So, it seems, you should not revise your beliefs.

But Sam's case is not like either of these: it seems that Sam *can* avoid being misled. He can point to a particular belief of his that is, he thinks, unsupported by his total evidence. He may even be able to identify a particular belief state that, he thinks, he should rationally adopt. (For example: supposing that Alex told him that his reasoning made him no more reliable than chance, Sam might think that the rational thing for him to do is to spread his credence evenly over the various candidates.) Unlike in the lottery case, for example, there is something in particular that Sam thinks he should believe—but he does not believe it.

If Level-Splitting is right, and extreme cases of epistemic akrasia can be rational in Sleepy Detective, there is nothing wrong with Sam's concluding that his evidence is misleading in this way. All Sam needed to get this reasoning started was high confidence in both "P" and "my evidence doesn't support P", plus the setup of the case: that E bears strongly on P. According to Level-Splitting, being in this kind of akratic state in situations like Sam's can be perfectly rational. But there *is* something wrong with Sam's concluding that his evidence is misleading in this case. This suggests that there is something wrong with Level-Splitting: specifically, there is something wrong with its denial of the Non-Akrasia Constraint.

So far we have seen two immediate counterintuitive consequences of Sam's reasoning from his akratic state in Sleepy Detective. Another way to bring out the oddness of Sam's belief state is to consider what it would look like to act on his beliefs; the irrationality of an agent's beliefs is often brought into sharper focus if we consider what happens when she puts her money where her mouth is. Perhaps if Sam is very confident that P is true, he will be willing to make a wager:

Sam: I'd bet that it's Lucy. I'll give you 9:1 odds.

Alex: But you were so sleepy when you were working last night! How can you be so sure that the evidence supports her guilt?

Sam: Oh, I'm not. Since you told me that I'm so bad at evaluating evidence when I'm tired, I doubt the evidence supports Lucy's guilt much at all. If I were to bet on what the evidence supported, I might give you 1:9 odds that it's Lucy, but certainly not 9:1.

Alex: So why are you offering 9:1 odds?

Sam: Well, I really *shouldn't* be offering such strong odds. I shouldn't be so confident
      that she's the thief: the evidence isn't in my favor. But on the other hand, she *is*
      the thief! That's what we're betting on, right?

Sam's behavior here illustrates the irrationality of believing as he does. It seems patently irrational to treat a bet about P and a bet about whether one's evidence supports P as completely separate. But if Level-Splitting is right, this is just how a rational person will behave in Sam's situation.[22]

And if Sam can act on his akratic beliefs, his odd-looking behavior won't be restricted to betting. Should Sam tell newspaper reporters that Lucy is the thief? Recommend that the cops raid her apartment? Send her to jail? If he is rationally highly confident that she is guilty, it seems that all of these actions could be warranted. But if Sam is asked to justify or explain his behavior, he will be at a loss: as far as he thinks, his evidence doesn't support any of it. In fact, he might think that he *shouldn't* be acting in these ways.[23]

Level-Splitting is thus committed to counterintuitive results beyond the initial strangeness of believing both "P" and "my evidence doesn't support P" in Sleepy Detective. It also seems to license patently bad reasoning and irrational action.

In the next section, I will look at some further drawbacks to the Level-Splitting view—among them, a different sort of bad reasoning—resulting from attempts to reconcile a Level-Splitting verdict in Sleepy Detective with intuitive verdicts in related cases.

4. Problems with Extending Level-Splitting

So far we have seen that Level-Splitting gives a certain verdict in cases like Sleepy Detective: Level-Splitting says that in such cases, one's confidence in P should be the level of confidence supported by one's first-order evidence alone, and one's confidence in "my evidence doesn't support P" should be the level of confidence supported by one's higher-order evidence alone. (Recall Weatherson's claim: "by hypothesis, both conjuncts are well-supported by the evidence.") But what should Level-Splitting say about other cases? Should epistemic levels always operate separately? Or should higher-order evidence rationally affect one's first-order attitudes under some circumstances (though not those in Sleepy Detective)? In this section I'll look at some possible ways one might extend the Level-Splitting view, and suggest that none of the options looks good.

4.1 No Interaction, Ever

A very strong version of Level-Splitting, on which higher-order evidence is *never* relevant to first-order attitudes, would be a non-starter. To see why, suppose that Sam and Alex are working on another case. All of the relevant evidence is down at the station; Sam is on his way in to work, and plans to evaluate the evidence when he gets there in order to determine which of the suspects is guilty. But Alex gets to the station first. Before Sam has seen any of the relevant evidence, Alex calls him up and tells him:

Alex: I've figured it out. All of the evidence points to Veronica's guilt!

Sam: Well, I don't care what the evidence points to. Let me see the evidence, so I can figure out who did it.

Alex: Veronica did it.

Sam: You could have said so in the first place! I thought you were just telling me what the evidence supported.

Sam's response here is irrational. When Alex tells him what the evidence supports, Sam should become confident that Veronica is guilty. But a strong, no-interaction-ever Level-Splitting view would not allow this. So Level-Splitters should not—and, I think, would not want to—adopt this extreme view.

4.2 Interaction Only in the Absence of Relevant First-Order Evidence

A better Level-Splitting view will need to allow higher-order evidence to rationally affect Sam's first-order beliefs in situations like the one described above, while denying that higher-order evidence rationally affects first-order beliefs in situations like Sleepy Detective. What is the difference between the two types of cases? One attractive hypothesis is that in Sleepy Detective, Sam *already has* the relevant first-order evidence. But in the case described in section 4.1, he does not. And this distinction is, plausibly, a significant one. Higher-order evidence is evidence about what a body of evidence supports. A Level-Splitter might argue that when higher-order evidence just bears on *your* evidence—that is, a body of evidence that you already have—it has no rational effect. But when higher-order evidence bears on a body of evidence that you *don't* have, it is rationally significant.

 If a Level-Splitter decides to go this way, one thing she might have in mind is something like this principle:

> **Proxy:** Higher-order evidence is relevant to first-order beliefs only insofar as it serves as a proxy for first-order evidence.

Proxy isn't explicitly endorsed by any of the people quoted in Section 2; their discussions of higher-order evidence are generally restricted to cases where one's higher-order evidence bears on first-order evidence that one already has, so the issue doesn't arise. But the idea behind Proxy is a fairly intuitive one, and could easily be adopted by any of the people quoted above. So before moving on, I will briefly look at Proxy, how it might motivate Level-Splitting, and how it might run into problems; this will give us a better idea of how higher-order evidence should work.

 First, let's see how Proxy might deliver the Level-Splitting verdict in the case above, where Alex tells Sam about evidence that he hasn't seen yet. When Alex tells Sam about the evidence she has seen, she is telling him something about what kind of evidence is out there. Since Sam knows a lot about detective work, it's possible that Alex's testimony tells him something about the *specific* first-order evidence that she saw. So when Alex tells Sam that all of the evidence points to Veronica's guilt, perhaps Sam should increase his confidence in the long disjunction of

first-order evidential propositions that would point to Veronica. (That is: that Alex has a gun with Veronica's fingerprints on it, *or* a letter in which Veronica describes her plans to an accomplice, *or...*) Since the truth of any of these disjuncts would itself support P, in becoming confident of the disjunction, Sam should also become more confident of P.

Second, Proxy could deliver the Level-Splitting verdict in cases where one learns that *there is* evidence for P, without identifying *what* that evidence is. For example, suppose that Sam, who knows nothing about physics, befriends a physicist. The physicist tells him that all of the evidence in her field strongly supports the proposition that the universe is expanding. Sam has no idea what such evidence would look like. (In fact, we can imagine that Sam knows so little about physics that for any candidate evidential proposition he can think of, he doesn't know whether it would count in favor of or against the first-order proposition in question.) Surely, in this case, Sam should become more confident that the universe is expanding. And Proxy can predict this.

So if a Level-Splitter endorses Proxy, she can maintain that higher-order evidence is relevant to first-order beliefs in some cases. She can also explain why, on her view, higher-order evidence is *not* relevant to Sam's first-order belief in Sleepy Detective. In Sleepy Detective, Sam already had all of the relevant first-order evidence. So it is implausible that his higher-order evidence could serve as a proxy for first-order evidence in either of the ways mentioned above. Sam's higher-order evidence should not increase his confidence in a long disjunction of first-order evidential propositions; he already knows which photos, letters, etc. compose the first-order evidence. And Sam's higher-order evidence should not increase his confidence that there is some first-order evidence at the police station that supports Veronica's guilt. That's because Sam already knows exactly what first-order evidence there is.

If higher-order evidence only affects first-order beliefs by Proxy, it will have no effect in cases where one already has the first-order evidence. In those cases, higher-order evidence will be *screened off*.

These observations suggest that a more plausible version of Level-Splitting, motivated by Proxy, should endorse something like the following prediction:

> Higher-order evidence should rationally affect one's first-order beliefs only insofar as one does not also possess the relevant first-order evidence.

When one has both first-order and higher-order evidence, a Level-Splitter can claim that first-order evidence should affect one's first-order beliefs, and higher-order evidence should affect one's beliefs about what one's evidence supports.[24] But she is also free to argue that, in these cases, higher-order evidence should *not* affect one's first-order beliefs. Proxy, then, seems to give Level-Splitters a tidy and intuitive way to distinguish between cases where higher-order evidence should rationally affect one's first-order beliefs and cases in which it should not.

But there is good reason to doubt that Proxy gives the whole story about how higher-order evidence works. Compare the following two cases:

> **Case 1:** Sam is working to find a serial killer. He has his evidence, E, set out in front of him, and he has carefully looked through all of it. E supports the proposition M: that

Marie is the killer. Sam has not yet finished his analysis of E and has not reached a conclusion about what E supports. Nevertheless, Sam just happens to believe—for no particular reason—that Marie is the killer.

Case 2: Sam and Alex are working together to find a serial killer. They have their evidence, E, set out in front of them, and they have both carefully looked through all of it. E supports the proposition M: that Marie is the killer. Sam has not yet finished his analysis of E and has not reached a conclusion about what E supports. But Alex finishes her analysis first, and exclaims, "I've got it! The evidence points to Marie." On the basis of Alex's testimony, Sam comes to believe that Marie is the killer.

These two cases have some things in common. In both cases, Sam's belief that M is supported by his first-order evidence, E. So there may be a sense in which Sam is required to believe M in both cases. But there also seems to be an important asymmetry between Case 1 and Case 2. In particular, Sam's belief that M seems much better justified in Case 2 than in Case 1. One way to put the point is like this: while Sam's belief that M is *propositionally* justified in both cases, it is only *doxastically* justified in Case 2.[25]

What could explain this asymmetry? Doxastic justification requires basing one's beliefs on evidentially relevant factors. So if Sam's belief in M is doxastically justified in Case 2, it must be because the basis for his belief in M—his higher-order evidence from Alex—is evidentially relevant to M. But Proxy does not predict this. After all, in Case 2, as in Sleepy Detective, Sam already knows what his first-order evidence is. Proxy says that when one has the relevant first-order evidence, higher-order evidence is not at all evidentially relevant to one's first-order beliefs. So if we want to capture the asymmetry between Case 1 and Case 2, we must allow that higher-order evidence can rationally affect first-order beliefs in ways other than by Proxy. A view on which Proxy is the *only* connection between epistemic levels will be missing something.

Perhaps Proxy points to *one* way in which higher-order evidence rationally affects first-order beliefs. But Case 2 suggests that this cannot be the *only* way. Rather, it seems that higher-order evidence must also affect first-order beliefs by bearing on evidential support relations: it is evidence about what some body of evidence supports.

Case 2, as a counterexample to Proxy, also spells trouble for Level-Splitting more generally. That's because any Level-Splitter who tries to accommodate this counterexample will have a hard time maintaining her verdict in Sleepy Detective. In Case 2, it seemed plausible that Sam's higher-order evidence should rationally bear on his first-order beliefs by way of bearing on evidential support relations. In Sleepy Detective, Sam's higher-order evidence seems to work in the same way, by bearing on the evidential support relation between E and P. So it's natural to think that Sam's higher-order evidence should have parallel effects in the two cases: if it rationally affects Sam's first-order beliefs in Case 2, it should do so in Sleepy Detective. Level-Splitters are therefore left with two options. They may either deny that there is any asymmetry between Case 1 and Case 2, or find a significant

disanalogy between Case 2 and Sleepy Detective to justify treating the two cases differently. Neither option looks promising.

4.3 Interaction in the Long Run

In this section we have so far looked at two questions about how the Level-Splitting view should be extended. We have seen that a plausible Level-Splitting view should allow higher-order evidence to affect first-order attitudes in some cases, but it turned out that one particularly tidy way of dividing these cases is unlikely to work.

A final question to consider is how Level-Splitting should deal with cases like Sleepy Detective if they happen again and again. Answering this question will be somewhat complicated, but looking at how a Level-Splitter might deal with a long string of cases will help further bring out what is wrong with Level-Splitting views.

Recall that Level-Splitting says that in cases like Sleepy Detective, one should become confident in P to the level supported by one's first-order evidence alone, and one should become confident in "My evidence doesn't support P" to the level supported by one's higher-order evidence alone. So in a single case like Sleepy Detective, Sam should become highly confident in each of those propositions.

What if Sam finds himself in cases like Sleepy Detective many times in a row? (We can imagine this happening in different ways: perhaps on the same night, Sam evaluated dozens of cases before talking to Alex on the phone; alternatively, several cases like this could occur sequentially.) If this happens, Sam will acquire what will seem to him like a surprisingly good track record: he will have high confidence in lots of true propositions, despite being quite unreliable at rationally assessing his evidence about those propositions. So Sam might reason along the following lines: "I believe P, Q, and R, and furthermore, P, Q, and R are true. Yet I formed my beliefs that P, Q, and R irrationally; I was rationally impaired (drunk, sleepy, etc.) when I did the calculations. It's highly unlikely that I would get all of these answers right, given that I wasn't rational in arriving at them."

According to Level-Splitting, what should Sam think has happened? I will consider two possible ways a Level-Splitter could answer this question, and suggest that either way raises new challenges for the view.

The first possible answer is that, even after the end of a long series of cases, Sam should remain epistemically akratic. That is, he should maintain his high confidence in P, Q, R, etc., and he should maintain his high confidence that his evidence doesn't support P, Q, R, etc. After a very long string of cases, how should Sam explain his success? Perhaps he should think he is *extremely* lucky. Strange things sometimes happen, and Sam could think that his long string of epistemic successes is one of them. Or perhaps Sam could increase confidence in some other hypothesis—one that would explain why he keeps getting things right, despite his unreliability at evaluating his evidence. If Sam goes this route, perhaps he should think that he has supernatural powers, or that a guardian angel is guiding him to the truth.

But none of these hypotheses seems like the best explanation for Sam's situation. After an apparently fair coin lands heads a hundred times in a row, you should not end up with high confidence that it's all a lucky coincidence, or that there is a heads-loving coin fairy. Instead, you should reduce your confidence that the coin

was ever fair in the first place. Analogously, the best explanation for Sam's situation is that he was rational all along: as he gets things right again and again, Sam should wonder if maybe he is better at evaluating his evidence than Alex suggested, and perhaps his evidence really supported his beliefs all along. So if Level-Splitters say that Sam should remain akratic at the end of a long sequence like this, they render him unable to rationally draw the conclusion that best explains his situation.

A second possibility for Level-Splitters is to say that, by the end of a long sequence of cases, Sam should no longer be epistemically akratic. The argument might go something like this. As Sam acquires a track record of true beliefs about P, Q, R, etc., he might take this track record to be evidence for his own reliability. The best explanation for reliability in reaching true beliefs is reliability in assessing one's evidence. So after a long enough string of successes, Sam should conclude that he was never seriously rationally impaired in the first place. That is, he should conclude that his evidence *did* support P, Q, and R. So he should not be akratic.

In adopting this strategy, Level-Splitters can claim that Sam *should*, in fact, draw the conclusion that best explains his situation. But the process by which Sam can come to this conclusion brings problems of its own. Sam's method of concluding that he was rational is an instance of *bootstrapping*—assessing the reliability of a belief process by using the process itself, without employing any independent checks. Bootstrapping in any context is notoriously suspicious.[26] In fact, Christensen raises the possibility of bootstrapping as an argument *against* the rationality of being epistemically akratic in a certain way: being certain of some logical theorem, but also doubting that one is perfectly rational.[27] If Level-Splitters say that Sam should bootstrap, their first challenge is to provide an explanation for why bootstrapping is an acceptable way to reason.

Of course, there are also bootstrapping apologists, and a Level-Splitter could always adopt that strategy too. For example, she might point out that, on her view, Sam can only rationally bootstrap if he has in fact evaluated his first-order evidence correctly in a large number of cases—so he will at least be prevented from bootstrapping to a false conclusion.[28]

But even if this response is satisfactory, Level-Splitters must meet a second, and more serious, challenge in defending the claim that Sam should bootstrap. The most natural way to get the bootstrapping reasoning off the ground is to note that truly believing P when one's evidence does not support it is unlikely or unexplained. If we need an explanation for being in such a state again and again, presumably it is because there is a kind of rational tension involved in being in this state *without* an explanation. Otherwise it would be perfectly rational to remain akratic even in a large number of cases. The thought behind this reasoning seems to rely on level-*bridging*: the idea that there is some kind of rational connection between epistemic levels. But level-bridging appears to be inconsistent with Level-Splitting: if there is a rational connection between epistemic levels, why doesn't higher-order evidence rationally affect our first-order beliefs in cases like Sleepy Detective?

Any level-connection that would explain why bootstrapping is rational here threatens to undermine the Level-Splitting verdict in a single case. And the reasoning that explains the Level-Splitting verdict in a single case (for example, Proxy)

seems unable to help justify bootstrapping. So if Level-Splitters want to say that Sam should no longer be akratic after a large number of cases like Sleepy Detective, they face the challenge of reconciling these two apparently incompatible claims: that in a single case, epistemic levels operate separately, but in the long run, they do not.

The bootstrapping argument is therefore one more instance of a general challenge for Level-Splitting. A Level-Splitter must maintain her verdict in a single case like Sleepy Detective, but should also allow that in certain similar situations—situations where Sam does not have the first-order evidence, for instance, and situations where the conditions in Sleepy Detective come about again and again—there is a rational connection between one's first-order beliefs and one's beliefs about what one's evidence supports. If Level-Splitters cannot tell a story about what kind of level-connection might support both judgments, they face a dilemma: they must either embrace absurdly strong Level-Splitting verdicts in a variety of situations, or give up their verdict in Sleepy Detective.

5. Bridge Principles, Reasoning, and Action

I've argued against having high confidence in "P, but my evidence doesn't support P" on the grounds that being in an akratic state like this, in cases like Sleepy Detective, licenses bad reasoning and irrational action. In order for these arguments to go through, I assumed that both "P" and "my evidence doesn't support P" were available to an akratic agent like Sam for the purposes of rational reasoning and action. These assumptions could be thought of as bridge principles, perhaps looking something like this:

> **Reasoning Bridge Principle:** It is rational to reason from any proposition, P, just in case one is rationally highly confident in P.[29]

> **Action Bridge Principle:** It is rational to act on any proposition, P, just in case acting on P maximizes expected value.

A Level-Splitter might then object: why hold these bridge principles? Without them, we need not say that epistemic akrasia licenses bad reasoning or irrational action. So in order to salvage the Level-Splitting view, the objection goes, we can just reject the bridge principles.

There are several ways one might fill in the details of this objection. For instance, one might argue that it is irrational to reason (or act) using beliefs that do not amount to knowledge (in this case, "my evidence doesn't support P").[30] Alternatively, one might argue that it is irrational to reason (or bet) using beliefs that one believes to be unjustified (in this case, "P").[31] Either way, an epistemically akratic agent like Sam wouldn't be able to rationally reason or act on both "P" and "my evidence doesn't support P". I will respond to this objection here.

First, we should note that giving up the Reasoning and Action Bridge Principles is not an obvious or easy move to make. For one, the Level-Splitting view and the motivations that count in its favor are at least prima facie independent of these

bridge principles. So denying the bridge principles is a substantial further claim, and would require additional argument.

And there is reason to think that this additional argument will be a tough sell: the bridge principles (or principles like them, at any rate) have considerable intuitive appeal. It's plausible that rational belief is closely connected with rational action, and that rational patterns of reasoning are determined by what one's evidence supports. And in order to block all of the apparently problematic consequences of epistemic akrasia, Level-Splitters will need to place significant restrictions on the bridge principles. (In order to say, for example, that Sam should not tell the police to arrest Lucy, Level-Splitters will have to go farther than denying that one can rationally place bets or make esoteric arguments.) These connections are, plausibly, part of what makes epistemic rationality worth theorizing about in the first place: looking at what we should believe helps us figure out what we should do. Those who are interested in epistemic rationality for these reasons should be especially hesitant to adopt a view on which the bridge principles do not hold.

Second, even if we give up the Reasoning and Action Bridge Principles, the arguments that appealed to these principles still have some bearing. To see why, let's just focus on Sam's argument that his evidence is misleading. The problem there seemed to be that Level-Splitting allowed Sam to irrationally draw a certain conclusion. If Level-Splitting denies the Reasoning Bridge Principle, the objection suggests, this problem would be avoided. But even if Sam cannot rationally draw the conclusion—for example, because he is prohibited from consciously going through the relevant argument, or because the justification for the argument's premises ("P" and "E supports low confidence in P") is not "transmitted" to the conclusion—the argument still shows that something is wrong with Level-Splitting. "P" and "E supports low confidence in P", together, *entail* the conclusion that E is misleading with respect to P. So regardless of whether or not Sam can use this conclusion to argue, act, or form new beliefs, his evidence supports it. That seems troubling enough. If it is implausible that Sam should believe that E is misleading, it is also—and for the same reasons—implausible that his evidence should support it.

The same goes for the other examples of bad reasoning or irrational action. The examples of bad reasoning show that, if Level-Splitting is right, in cases like Sleepy Detective our evidence supports odd and counterintuitive conclusions beyond "P, but my evidence doesn't support P". The examples of irrational action bring out what is intuitively problematic about representing the world in the way that, according to Level-Splitting, is supported by our evidence. Even without the bridge principles, both kinds of argument give us reason to worry about the evidential support relations endorsed by Level-Splitting. And since Level-Splitting is a view about rational belief, a reason to worry about Level-Splitting's evidential support relations just is a reason to worry about the view itself.

6. Possible Exceptions?

Sleepy Detective is a paradigmatic example of the type of case usually discussed in the higher-order evidence debate, and so far we have seen several problems with

the view that epistemic akrasia is rational in that case. But recent literature has suggested that under certain special conditions, higher-order evidence might work differently from how it does in cases like Sleepy Detective. And some authors—even some authors who defend Conciliatory views about cases like Sleepy Detective—have argued that under those special conditions, it is rational to be epistemically akratic.[32] What does this mean for the Non-Akrasia Constraint, and the arguments we have seen so far in its support?

I will close by considering an example of this type of special case, which I'll call "Dartboard". As we will see, epistemic akrasia seems rational in Dartboard only if we make the (controversial) assumption that we can be rationally uncertain about what our evidence is. One might wish to deny that assumption for independent reasons, in which case Dartboard will not seem like much of a worry. But for those who find the assumption plausible, admitting epistemic akrasia in Dartboard looks like serious cause for concern. To answer this concern, it will be worth looking at the relationship between Dartboard and Sleepy Detective.

In this section I will argue that accepting a pro-akrasia verdict in Dartboard does not commit us to Level-Splitting. On the contrary, examining why epistemic akrasia might be rational in this type of special case will help us see why Level-Splitting is wrong. I will also suggest that, surprisingly enough, those who accept pro-akrasia verdicts in cases like Dartboard can still argue against Level-Splitting by appealing to some of the main anti-akrasia arguments that I have offered here.[33]

I will focus on the following scenario, adapted from Williamson;[34] it is similar to Williamson's puzzle of the unmarked clock:[35]

Dartboard: You have a large, blank dartboard. When you throw a dart at the board, it can only land at grid points, which are spaced one inch apart along the horizontal and vertical axes. (It can only land at grid points because the dartboard is magnetic, and it's only magnetized at those points.) Although you are pretty good at picking out where the dart has landed, you are rationally highly confident that your discrimination is not perfect: in particular, you are confident that when you judge where the dart has landed, you might mistake its position for one of the points an inch away (i.e. directly above, below, to the left, or to the right). You are also confident that, wherever the dart lands, you will know that it has *not* landed at any point farther away than one of those four. You throw a dart, and it lands on a point somewhere close to the middle of the board.

We can illustrate Dartboard like this (suppose the board is unmarked and much larger, so that it is not obvious where the dart is):

Let's assume the dart landed at <3,3>. Given that the dart landed at <3,3>, what should you believe about where the dart landed?

In this case, many authors find it intuitively plausible that, for any particular point, you should not be certain that that point is where the dart landed. After all, it's a large blank board, and you have good reason to think that you might confuse adjacent points with one another. Nevertheless, you should be able to narrow down the possibilities significantly: your discrimination may not be perfect, but it's still pretty good. So let's suppose that when the dart lands at <3,3>, you should be highly confident in the proposition that it landed at either <3,2>, <2,3>, <3,3>, <4,3>, or <3,4>—so, you can rationally rule out every point except for those five. (If this margin for error seems implausible, expand or shrink your imagined dartboard as necessary.) Williamson agrees with this verdict, and supposes further that your credence should be equally distributed over <3,2>, <2,3>, <3,3>, <4,3>, and <3,4>. So, for each of those five points, you should have .2 credence that the dart landed at that point.

Now consider the proposition Ring: that the dart landed on one of <3,2>, <2,3>, <4,3>, or <3,4>. (That is, that it landed on one of the points in the ring around <3,3>.) Your credence in Ring should be .8. But that level of confidence in Ring is *only* rational if the dart actually landed on <3,3>. If the dart had landed on some other point, the rational credence distribution would be centered on that point instead of <3,3>, and your rational credence in Ring would be *lower* than .8. In particular, if the dart landed at any of <3,2>, <2,3>, <4,3>, or <3,4>, it would be rational for you to have .2 credence in Ring. Suppose you have reflected on your situation, and you know what would be rational to believe in these various evidential situations. So you should be .8 confident of Ring, but also .8 confident that your credence in Ring should be .2. This means you should be epistemically akratic: you should be highly confident of both "Ring" and "my evidence supports ~Ring".[36]

How should we think about this case, given that we have already seen how bad epistemic akrasia is in Sleepy Detective? One possibility is to reject Williamson's verdict. Perhaps we could do this by moving to an extreme view on which we should never be rationally uncertain of what our evidence is or what it supports; this would rule out all kinds of epistemic akrasia as rational, including that motivated by Dartboard. But if our main interest is to preserve an anti-akrasia verdict in Sleepy Detective, such extreme measures are not necessary. For the remainder of this section, I will adopt the assumption that Williamson's verdict in Dartboard-style cases is right, and look at how this verdict squares with an anti-akrasia view about Sleepy Detective.

One difference we should notice right off the bat is that the akratic state required in Dartboard is somewhat different from the akratic state that Level-Splitters require in Sleepy Detective. In Sleepy Detective, Level-Splitters say that you can rationally (a) be certain of what your evidence is, (b) be highly confident that it supports ~P, and (c) have high confidence in P anyway. So you are akratic because you are uncertain (or have a false belief) about what E, your evidence, *supports*. In Dartboard, however, you are akratic because (in part) you are

uncertain about what your evidence *is*. This is compatible with your being highly confident, or even certain, of the truth about the evidential support relations. So while the two akratic states look similar, they come about through different kinds of uncertainty.[37]

The second difference between the cases—and the one I will focus on at greater length—is in how we should expect the evidence to vary with the truth of the propositions it supports: Lucy's guilt in Sleepy Detective, and Ring in Dartboard.[38] In cases like Sleepy Detective, our evidence is usually "truth-guiding" with respect to propositions about the identity of the guilty suspect (and most other propositions, too). By this I mean simply that the evidence usually points to the truth: when it justifies high confidence in a proposition, that proposition is usually true, and when it justifies low confidence in a proposition, that proposition is usually false. If a detective's first-order evidence points to a particular suspect, that suspect is usually guilty. If it points away from a particular suspect, that suspect is usually innocent.

This feature of Sleepy Detective is by no means exceptional. In most cases, we should expect our evidence to be truth-guiding: otherwise, why should we ever believe what it supports? Moreover, this feature of cases like Sleepy Detective is something that should be available to an agent like Sam: Sam should know, even before looking at his evidence, that evidence of this type usually supports the truth about which suspect is guilty.

With this in mind, let's think about what happens if Sam becomes epistemically akratic in Sleepy Detective, as Level-Splitting suggests. If he is akratic, he cannot continue to believe that his evidence is truth-guiding. Although he will be highly confident that the evidence supports Lucy's *innocence*—a condition that would normally go along with her actually *being* innocent—he will continue to be highly confident of her *guilt*. Epistemic akrasia seems so odd in this case, in part, because it involves ignoring the reasonable background expectation that one's evidence is truth-guiding.

In Dartboard, however, the evidence is *not* truth-guiding, at least with respect to propositions like Ring. Instead, it is *falsity*-guiding. It supports high confidence in Ring when Ring is false—that is, when the dart landed at $<3,3>$. And it supports low confidence in Ring when Ring is true—that is, when the dart landed at $<3,2>$, $<2,3>$, $<4,3>$, or $<3,4>$. This is an unusual feature of Dartboard. And it is only because of this unusual feature that epistemic akrasia seems rational in Dartboard. You should think that you should have low confidence in Ring precisely *because* you should think Ring is probably true—and because your evidence is falsity-guiding with respect to Ring. Epistemic akrasia is rational precisely because we should take into account background expectations about whether the evidence is likely to be truth-guiding or falsity-guiding.

The reasoning that we just went through has a strange result in Dartboard: it justifies epistemic akrasia. But if we apply that same reasoning to a more standard case like Sleepy Detective, where our evidence is truth-guiding, we get the standard Conciliatory verdict. Returning to that case, suppose that Sam starts off rationally expecting his first-order evidence to be truth-guiding. Suppose he also has higher-order evidence (Alex's testimony) suggesting that his likelihood of interpreting the

first-order evidence correctly is no better than chance. Even after seeing the first-order evidence, then, Sam should remain highly uncertain about what it supports. And because he still expects his evidence to be truth-guiding, his uncertainty about what the evidence supports should bring with it uncertainty about which suspect is guilty. So even after seeing the evidence, Sam should be highly uncertain about which suspect is guilty. For any particular suspect, including Lucy, he should end up with low credence that that suspect is guilty. And this is exactly what the Conciliatory view recommends.

Now let's return to some of the problems for epistemic akrasia, and see how these arguments work in Dartboard. First, let's consider Sam's conclusion that his evidence is *misleading* in Sleepy Detective (from section 3, above). I argued above that it is irrational for Sam to believe that his evidence is misleading, in part, because it seems that he can identify something he could do to avoid being misled. In other cases, however, believing that one has misleading evidence might not be so bad. Specifically, it did not seem at all problematic in cases like the lottery, where there is nothing one can do to avoid being misled. In Dartboard, it does seem that you should think that your evidence is misleading. You should, after all, have .8 credence in Ring, but you should also be highly confident that your credence in Ring should be .2. But if we look more closely at the details of Dartboard, we can see that it might be best treated as a case like the lottery, where one cannot avoid being misled.

Here is why: in Dartboard, you should be .8 confident that the dart landed on one of <3,2>, <2,3>, <4,3>, or <3,4>, but you have no idea which. Each of those evidential situations rationalizes a different credence distribution. And although those different credence distributions each assign .2 credence to Ring, they disagree about nearly everything else. For example, if the dart landed on <3,2>, you should have .2 credence in Ring because you should have .2 credence that the dart landed on <3,2>. But if the dart landed on <3,4>, you should have .2 credence in Ring because you should have .2 credence in <3,4>—and 0 credence in <3.2>! So although you should be .8 confident that you should have .2 confidence in Ring, there is no particular credence distribution that you think you should adopt. While you should think your evidence is misleading, this belief nevertheless seems stable.

Second, let's return to the bootstrapping argument. I argued that if Sam becomes epistemically akratic in a long string of cases like Sleepy Detective, he will find himself in an unusual epistemic situation—one that requires explanation. The best explanation for that situation is that he was never seriously rationally impaired in the first place. So, I argued, Sam should bootstrap.

The bootstrapping argument relies on the fact that in cases like Sleepy Detective, we rationally expect our evidence to be truth-guiding. This is why believing irrationally is usually unreliable, why finding oneself epistemically akratic is un-usual, and why believing rationally is a good explanation for one's having true beliefs. But when one's evidence is *falsity*-guiding, as it is in Dartboard, things may be different. Finding oneself with true-but-irrational beliefs is to be expected in cases like Dartboard. And because one's evidence is not truth-guiding, believing rationally is not a good explanation for having true beliefs about propositions

like Ring. In Dartboard, then, the bootstrapping argument does not get off the ground.

If this type of Conciliatory view is right, we cannot endorse the Non-Akrasia Constraint as a universal prohibition against a certain combination of attitudes. That's because it might sometimes be rational to have high confidence in both "P" and "my evidence doesn't support P". But contra Level-Splitting, this combination of attitudes is not made rational merely by our receiving misleading higher-order evidence. Instead, epistemic akrasia will only be rational in those odd cases where we expect our evidence to be falsity-guiding. The next task for this type of Conciliatory view, then, is to develop a more nuanced account of the interaction between epistemic levels: one that takes into account whether we rationally expect our evidence to point to the truth.[39]

Conclusion

I have argued that Level-Splitting views—those on which epistemic akrasia can be rational in cases like Sleepy Detective—incur substantial intuitive costs. This is good evidence for the Non-Akrasia Constraint: the problems for Level-Splitting suggest that at least in paradigmatic cases, it cannot be rational to have high confidence in both "P" and "my evidence doesn't support P".

Some important questions remain. First, it remains to be seen *how* higher-order defeat might work: what if, as Lasonen-Aarnio argues, there is no satisfactory, coherent way of spelling this out? And second, much more should be said about *when* there is higher-order defeat: when should higher-order evidence affect our first-order beliefs? And when, if ever, can epistemic akrasia be rational?

Though the discussion here doesn't settle these questions once and for all, we do now know something about how a good theory of higher-order evidence will answer them. First, in order to be extensionally adequate, a good theory of higher-order evidence should say that, in paradigmatic cases like Sleepy Detective, epistemic akrasia cannot be rational. That's because in those cases, epistemic akrasia leads to bad reasoning and irrational action. And we should take these problems seriously even in the absence of general principles: it is difficult to imagine even a rough picture of the relationship between epistemic levels that would vindicate a pro-akrasia verdict in these paradigmatic cases without committing us to absurd consequences.

Second, we now know something about *why* higher-order evidence should affect our first-order attitudes in paradigmatic cases of misleading higher-order evidence. That is: in normal cases, we rationally expect our evidence to be truth-guiding. And when we expect our evidence to be truth-guiding, what we should believe about the world should line up with what we believe our evidence supports. We should not have high confidence in both "P" and "my evidence doesn't support P".[40]

Notes

[1] Some authors take the irrationality of epistemic akrasia as a premise. See, e.g., Smithies [2012] and Titelbaum [ms]. Feldman [2005] defends a broadly anti-akratic view, and writes that "[o]ne wonders

what circumstances could make [epistemic akrasia] reasonable" (p. 108–9). Christensen [2007b] argues that it's not rational to be certain of logical theorems while doubting that you are perfectly reliable at proving these theorems (p. 18–19). Many arguments in the literature on disagreement and higher-order evidence rely on the idea that if it is rational for you to believe that a certain belief of yours is unjustified, you should reduce confidence in that belief. See, e.g., Christensen [2007a], [2010], and [2011]; and Elga [2007] and [ms]. Kolodny [2005] assumes that epistemic akrasia is a type of internal conflict, and suggests that one's evidence will never support being in such a state. Adler [2002] argues that it is psychologically impossible to believe P, be fully aware that one believes P, and also believe that one's belief in P is unsupported by one's evidence. Greco [forthcoming] argues that we should understand akratic belief states as "fragmented", and that this justifies our taking them to be irrational.

For variety's sake, I'll sometimes describe akratic states as beliefs or high confidence in conjunctions, and sometimes as combinations of attitudes. Strictly speaking, the Non-Akrasia Constraint should be understood simply as a prohibition on having high confidence in both "P" and "my evidence doesn't support P".

[2] See Kelly [2005] and Titelbaum [ms] for defenses of this type of view, applied specifically to the case of peer disagreement.

[3] See Feldman [2005] for a view like this; see also the recent literature on disagreement and higher-order evidence by, e.g., Christensen and Elga. Kelly [2007]'s "Total Evidence" view could be read as a compromise between Steadfast and Conciliatory views that still respects the Non-Akrasia Constraint. So, even among broadly Conciliatory views, just *how much* Sam should decrease confidence is up for debate. Sleepy Detective is meant to parallel cases discussed by all of these authors regarding higher-order evidence.

[4] See Williamson [notes a], p. 6, and [notes b], p. 36–42, for discussion of this kind of case. In his [notes b], fn 19, Williamson explicitly discusses cases where one's lack of higher-order knowledge comes about because of doubts that one's deduction is free from "inferential error".

[5] Thanks to an anonymous *Noûs* referee for helpful suggestions regarding this point.

[6] Williamson [notes b], p. 2.

[7] See Williamson [2000], p. 229. Williamson describes a case in which P is 80% probable on one's evidence, but it is 80% probable that P is only 20% probable. This example is a useful illustration of one of the many ways in which Williamson's view allows one's rational confidence in "P" and "my evidence supports P" to come apart quite widely when one has reason to be uncertain about one's evidence. This particular case deserves further attention, and I will discuss it in detail in Section 6. For present purposes, however, I will focus on cases like Sleepy Detective.

[8] This is not to say that there is nothing wrong, epistemically speaking, with someone who maintains confidence in P after receiving evidence that her belief in P is unjustified or unsupported by her evidence. Lasonen-Aarnio suggests that there is a separate epistemic norm, "reasonableness," that makes this kind of belief state criticisable too. See also her [2010] for more on this idea.

[9] Lasonen-Aarnio [forthcoming], p. 24.

[10] Wedgwood [2011], p. 21.

[11] Ibid.

[12] Ibid.

[13] Weatherson [ms], p. 15. Additionally, in his [2008], Weatherson tentatively endorses the thought that "[i]t is possible for S to have a justified but false belief that her belief in p is justified." (p. 22) If this is right, a natural consequence is that it could be rational to believe something like, "I should believe P", while simultaneously disbelieving or suspending judgment on P.

[14] Coates [2012], p. 122.

[15] Ibid., p. 12–13. Coates ([2012], p. 116) discusses the story of Huck Finn, who saves Jim, a runaway slave, while believing that he ought to turn him in to his "owner". The Huck Finn case adds an additional complication: Huck's belief about what he ought to do is not, according to Coates, epistemically justified. Weatherson's case is somewhat more straightforward, provided that his account of the Kantian student's epistemic and moral justification is right.

[16] Coates [2012], p. 113.

[17] Later in his [ms], Weatherson considers a weakened version of Level-Splitting, whereby HOE would fail to affect Sam's first-order beliefs only in cases where the first-order belief is the result of

a "basic inference", and where HOE is "relatively weak". (Weatherson [ms], p. 12.) "Relatively weak" higher-order evidence seems to mean evidence that reduces one to a precise state of neutrality over whether one's evidence supports P. I see no good explanation for why higher-order evidence should function in this strange way under only these rare, knife-edge conditions. Furthermore, if Weatherson's moral analogy works in the way it's supposed to, it seems to support the rationality of more strongly akratic epistemic states. I discuss this further in the next subsection.

[18] Hazlett [2012].

[19] See Smithies [2012] for further discussion of this point.

[20] A Level-Splitter might object to this point by denying that Sam is rationally permitted to reason from both of these propositions. I will address this objection in section 5.

[21] Thanks to an anonymous *Noûs* referee for suggesting that I consider this type of case.

[22] Coates discusses a similar betting argument, and suggests that it can be avoided by holding that agents in Sam's situation should not act on their akratic belief states. He writes, "it is a [practically] rational policy not to make bets one believes are irrational, since acting on this policy will likely maximize expected value in the long run". (Coates [2012], p. 121.) In other words, instead of following the rule "take bets that maximize expected value", we should follow something like "take bets that you believe will maximize expected value." Coates acknowledges that following this second rule will lead one to *not* maximize expected value in some particular cases; so, we would in fact do better by following a policy of maximizing expected value. In order to take Coates' response on board, we must either deny that it is rational to maximize expected value, or else posit two types of rational evaluation. For example, there might be one norm that tells us to follow the rules that would be best if followed, and one norm that tells us to follow the rules that would be best if we *tried* to follow them. (This second proposal is perhaps comparable to Lasonen-Aarnio's epistemic norm of "reasonableness"; see fn 8.) Coates' response deserves further discussion, but I will not pursue the suggestion here.

[23] Again, a Level-Splitter might object to this point by denying certain bridge principles between rational degrees of belief and action. I will return to this objection in section 5.

[24] Weatherson considers this position, which he calls "ESJ" (for "Evidence Screens Judgment") in his [ms]. Though he doesn't explicitly endorse the position, much of what he says is consistent with the thought that, at least as far as first-order beliefs go, first-order evidence screens off higher-order evidence.

[25] One might object that Sam's belief that M in Case 2 is not *fully* doxastically justified, since it is not based on his total relevant evidence. But even if we grant this, there still seems to be an asymmetry between the two cases: Sam's belief that M seems to be *somewhat* doxastically justified in Case 2, and is not *at all* doxastically justified in Case 1.

[26] See, e.g., Vogel [2000] and Cohen [2002].

[27] See Christensen [2007b], p. 18. Christensen's discussion combines the possibility of bootstrapping with the possibility of extreme luck; if what I say here is right, however, Level-Splitters should see these possibilities as very different. See also White [2009]. Elga [2007] makes a similar argument in the context of peer disagreement.

[28] Kornblith [2009] defends reliabilism on roughly these grounds. Kelly ([2007], esp. section 5.4) makes a similar argument defending his "Total Evidence" view of peer disagreement against Elga [2007]'s objection that Kelly's view sanctions bootstrapping.

[29] Of course, someone defending this principle should also require that one's reasoning preserve coherence, respect one's total evidence, and the like. Spelling out these details might best be thought of as giving norms of *good* reasoning; the bridge principle I discuss here is just supposed to determine which propositions one can rationally use in one's reasoning in the first place.

[30] Williamson, for example, might want to go this way. Since knowledge is the norm of both assertion and belief, on his view, it is plausible that the view might also prohibit one from reasoning from premises that are not known.

[31] Coates' discussion of the betting argument employs a version of this strategy. See fn 22.

[32] See, e.g., Christensen [2010] and Elga [forthcoming].

[33] Special thanks to Bernhard Salow for helpful discussion on this topic.

[34] See Williamson [2000], p. 229. Williamson presents the case abstractly; Dartboard is a slightly more realistic version with the same formal properties.

[35] The original clock case is from Williamson [notes a] and [notes b], and is also discussed in Christensen [2010] and Elga [forthcoming]. The only difference between Dartboard and the clock case is that in Dartboard, the evidence varies along two dimensions rather than one. I discuss Dartboard rather than the clock because, in a two-dimensional case, it is more plausible that there can be a large divergence between first-order and higher-order attitudes: that is, that it can be rational to be highly confident of both "P" and "my evidence supports ∼P", rather than the more moderate-sounding "P" and "my evidence supports slightly lower confidence in P than I actually have". That is, Dartboard seems to motivate "extreme" epistemic akrasia in the sense I discuss in section 2.5.

[36] You might object that this extreme result only comes about because we assumed that your credence should be distributed evenly over the five possibilities. In a more realistic case we should want to give <3,3>, the point where the dart actually landed, at least a bit more weight. It's true that the divergence between your credence in *Ring* and your credence that *your evidence supports ∼Ring* will get smaller and smaller as we put more weight on <3,3>, but the general problem will not go away so easily. In a three- or four-dimensional case, putting a lot of credal weight on the actual case becomes less plausible. So it would be hard to admit that Williamson's description of Dartboard-style cases is *ever* right without also admitting that there are some cases in which extreme epistemic akrasia is rational.

[37] See Titelbaum [ms] for further discussion of this distinction.

[38] See Elga [forthcoming] for a more condensed version of this point, which Elga brings up in defense of his New Rational Reflection principle.

[39] Elga's "New Rational Reflection" principle is one promising attempt to do just that. So, there is reason to be optimistic about the prospects for this type of Conciliatory view. See Elga [forthcoming].

[40] For helpful comments, suggestions, and questions, I am grateful to David Gray, Daniel Greco, Brian Hedden, Jack Marley-Payne, Katrina Przyjemski, Bernhard Salow, Miriam Schoenfield, Paulina Sliwa, Robert Stalnaker, Katia Vavova, Brian Weatherson, Stephen Yablo, and audiences at the MIT Epistemology Reading Group and the 2011 Columbia/NYU Graduate Conference. Special thanks to David Christensen and Roger White, for valuable conversations and feedback on many earlier drafts of this paper, and to an anonymous referee for *Noûs*, whose extensive written comments helped improve the paper greatly.

References

Adler, Jonathan. [2002] *Belief's Own Ethics*. MIT Press.

Christensen, David. [2007a] "Epistemology of Disagreement: the Good News". *The Philosophical Review* 116 (2): 187–217.

——— [2007b] "Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals". *Oxford Studies in Epistemology* 2: 3–31.

——— [2010] "Rational Reflection". *Philosophical Perspectives, volume 24: Epistemology* (Blackwell, Oxford) pp. 121–140.

——— [2011] "Disagreement, Question-Begging and Epistemic Self-Criticism". *Philosopher's Imprint* 1 (6).

Coates, Allen. [2012] "Rational Epistemic Akrasia". *American Philosophical Quarterly* 49 (2): 113–124.

Cohen, Stewart. [2002] "Basic Knowledge and the Problem of Easy Knowledge". *Philosophy and Phenomenological Research* 65 (2): 309–39.

Elga, Adam. "The Puzzle of the Unmarked Clock and the New Rational Reflection Principle." Forthcoming in *Philosophical Studies*.

——— [2007] "Reflection and Disagreement". *Noûs* 41 (3): 478–502.

——— [ms] "Lucky to be Rational."

Feldman, Richard. [2005] "Respecting the Evidence". *Philosophical Perspectives, volume 19: Epistemology* (Blackwell, Oxford) pp. 95–119.

Greco, Daniel. "A Puzzle About Epistemic Akrasia". Forthcoming in *Philosophical Studies*.

Hazlett, Allan. [2012] "Higher-Order Epistemic Attitudes and Intellectual Humility". *Episteme* 9 (3): 205–223.

Kelly, Thomas. [2010] "Peer Disagreement and Higher Order Evidence". In Richard Feldman and Ted Warfield (eds.) *Disagreement* (Oxford University Press), pp. 111–174.

——— [2005] "The Epistemic Significance of Disagreement". In John Hawthorne and Tamar Szabo Gendler (eds.) *Oxford Studies in Epistemology*, volume 1 (Oxford University Press), pp. 167–196.

Kornblith, Hilary. [2009] "A Reliabilist Solution to the Problem of Promiscuous Bootstrapping". *Analysis* 69 (2): 263–267.

Lasonen-Aarnio, Maria. "Higher-Order Evidence and the Limits of Defeat". Forthcoming in *Philosophy and Phenomenological Research*.

——— [2010] "Unreasonable Knowledge". *Philosophical Perspectives* 24 (1): 1–21.

Smithies, Declan. [2012] "Moore's Paradox and the Accessibility of Justification". *Philosophy and Phenomenological Research* 85 (2): 273–300.

Titelbaum, Michael. [ms] "In Defense of Right Reasons".

Vogel, Jonathan. [2000] "Reliabilism Leveled". *Journal of Philosophy* 97 (11): 602–623.

Weatherson, Brian. [ms] "Do Judgments Screen Evidence?"

——— [2008] "Deontology and Descartes' Demon". *Journal of Philosophy* 105 (9): 540–569.

Wedgwood, Ralph. [2011] "Justified Inference". *Synthese* 189 (2): 1–23.

White, Roger. [2009] "On Treating Oneself and Others as Thermometers". *Episteme* 6 (3): 233–250.

Williamson, Timothy. [2000] *Knowledge and its Limits*. Oxford University Press.

——— [notes a] "Improbable Knowing".

——— [notes b] "Very Improbable Knowing".