Formulating Independence[1]

1. Introduction

People often encounter evidence which bears directly on the reliability or expected accuracy of their thinking about some topic. This has come to be called "higher-order evidence".[2]

For example, suppose I have some evidence E, and come to have high confidence in hypothesis H on its basis. But then I get some evidence to the effect that I'm likely to do badly at assessing the way E bears on H. Perhaps E bears on H statistically, and I'm given evidence that I'm bad at statistical thinking. Or perhaps E is a set of CVs of male and female candidates, H is the hypothesis that a certain male candidate is a bit better than a certain female candidate, and I get evidence that I'm likely to overrate the CVs of males relative to those of females. Or perhaps E consists of gauge and dial readings in the small plane I'm flying over Alaska, H is the hypothesis that I have enough fuel to reach Sitka, and I realize that my altitude is over 13,000 feet, which I know means that my reasoning from E to H is likely affected by hypoxia. Or finally, perhaps E is a body of meteorological data that seems to me to support rain tomorrow, H is the hypothesis that it'll rain tomorrow, and I learn that my friend, another reliable meteorologist with the same data E, has predicted that it won't rain tomorrow.

In cases like these, it seems that part of what determines how confident I should end up being in H is some kind of assessment of my own reliability, or expected accuracy, in forming my initial credence in H. So, in general, the credence I end up with should be informed by what I'll call a "reliability-assessment".

The cases also illustrate the range of considerations that should inform this reliability assessment. They include factors that bear on my general competence, my psychological quirks, my current circumstances, and even, in the case of disagreement, generic evidence that I've screwed up somehow or another today. But there's one way of assessing my reliability that seems clearly irrational: I should not reason that since E strongly supports H, and I've come to believe H on the basis of E, that my thinking in the present instance is fine, despite my poor statistical skills/my implicit bias/my likely hypoxia/my friend's disagreement. I should not assess my reliability in this way, even if my reasoning from E to H happens, in this case, to be perfectly correct. The relevant reliability assessment needs to be *independent* of this sort of reasoning.

Nevertheless, it turns out to be difficult to formulate a plausible Independence principle. This paper aims to make some progress on this project, and also to illustrate some of the difficulties that it entails. We might start with a highly contrived example, devised to make it maximally easy to see what should, and what should not, be allowed to inform the relevant independent reliability

---

[2] The term 'higher-order evidence' has actually been given several non-equivalent understandings in the literature. Some take it as evidence bearing on evidential relations (which makes the term particularly appropriate). Others focus on evidence about the rationality of the person's thinking, or (as I will here) on evidence about the reliability or accuracy of the person's thinking. As Jonathan Vogel pointed out to me, it's not clear that the latter understandings are naturally captured by 'higher-order'. But the literature has focused on a common set of examples—examples of the sort I'll concentrate on here. So I'll go with the 'higher-order" terminology.

assessments. This will allow us to sketch an Independence principle, which can then be refined and tested by considering less straightforward cases.

> **Logic on Drugs**: Alicia is told two things by a source she rationally believes to be highly reliable:
> > (A) Karla was born in May if and only if Kayla wasn't; and
> > (B) Either Kayla, or Layla and Lola, the Lumpkin twins, were born in May.
> She rationally becomes extremely confident in A and B, thinks about their implications a bit, and rationally becomes highly confident—say, .99 confident—that:
> > (P) Karla wasn't born in May unless Layla Lumpkin was.
> Then Alicia learns that before she started to think about all these birthdays, someone slipped her a powerful drug. The drug distorts people's complex truth-functional reasoning about birthdays. It causes people who reach high credences from doing this sort of reasoning to favor incorrect conclusions in 30% of the time. Alicia has played with this drug before, at parties. When she forms high credences, she has a long history of forming them in wrong conclusions 30% of the time, in problems just like this one, even while feeling perfectly clear-headed. She reflects on all this, and becomes significantly less confident in P—say, she reduces her confidence to around .7.

I will assume that the credence she adopts is the one that's most rational for her, and that the credence she's most rational to adopt depends on an assessment of her likely reliability on this occasion. It's clear that this reliability-assessment should take into account some facts about her present situation: in particular, that she has ingested the logic-disrupting drug. But just as clearly, the reliability-assessment relevant to Alicia's final credence in P should not be informed by the following train of reasoning:

> 1. A
> 2. B
> So,   3. P
> 4. I came to believe P
> So,   5. The drug did not interfere with my reliability today!

This train of reasoning would seem to beg the question in some sense: it would dismiss the worry about a bit of Alicia's reasoning by relying on the very reasoning in question. Putting the point informally, it seems that the relevant reliability assessment must be *independent* of Alicia's reasoning from A and B to P.

But how might we make this informal statement more precise? In trying to make progress on this question, it will be useful to begin by restricting attention to a certain class of cases: ones where the agent begins by forming a *rational initial credence* in some proposition P based on a *well-defined bit of first-order evidence*, and then learns some higher-order evidence that bears on the reliability of the reasoning by which that first-order evidence led to her credence in P. I will also assume that the agents form *rational* independent reliability-assessments.

It will also be useful to work with a particular model of what *form* the reliability-assessments might take. I'll work here with a model (defended elsewhere[3]) on which they take the form of *hypothetical conditional credences*. In Alicia's case, we may think of her as focusing on the reliability of her own reasoning from her very high credence in A and B to her initial very high credence P. We can ask how likely Alicia should think it is that P is really true, conditional on the fact that that she,

---

[3] See Christensen (2016).

having ingested the drug, reached that very high credence in P on the basis of A and B. If this conditional credence is not to be affected by the sort of question-begging reasoning described above, it will have to be *independent* from Alicia's reasoning from A and B to P. Intuitively, we can see Alicia as stepping back from her confidence in P, and considering herself as a kind of measuring device, whose high credence in P serves as an indicator of how likely P is to be true. I take it that Alicia's independent hypothetical credence in P, given the information about her being drugged and about the results of her initial reasoning, should be about .7.

Here, then, is a very rough first pass at characterizing reliability estimates that are independent in the requisite way:

> **Independence, preliminary sketch:** When an agent has formed an initial credence c in P on the basis of first-order evidence E, and then gets some evidence that bears on the reliability of her reasoning from E to her credence in P, her final credence in P should reflect the Independent Hypothetical Credence (IHC) it would be rational for her to have in P: that is, the rational credence in P independent of her reasoning from E to her initial credence in P, but conditional on her having formed credence c in P on the basis of E, and on the reliability evidence the agent has about herself.[4]

In the present example, the agent reached very high initial credence, and the rational IRC for her was considerably lower. But it is important that this need not always be the case. Suppose that an agent is hoping for a nice hike tomorrow, and her direct assessment of the meteorological evidence resulted in credence .5 in rain. Then she gets good evidence that she tends to reach overly-optimistic opinions about the weather. In such a case, the agent's rational IHC in rain, given that she reached .5 initial credence in rain, and given the reliability evidence she now has about herself, would be higher than .5 (the amount by which it would be higher would depend on the specifics of the evidence about her optimism). So IHCs can be lower, or higher, or even equal to agents' initial credences.[5]

With this framework in hand, let us turn to making the intuitive idea more precise, and then to consider some cases where applying Independence will be more difficult.

2. Independent from "the agent's reasoning"?

The above description of Alicia's reliability-assessment requires that it be "independent from Alicia's reasoning from A and B to P." But what this means is hardly clear. Similar formulations in the disagreement literature include "independent of the dispute" and "independent of the disagreement." Less-common formulations talk about independence from the initial belief, and independence from the first-order "evidence" on which the initial belief was based, or independence from "the reasons" for the initial belief (which seems ambiguous between independence from the

---

[4] In this and subsequent formulations, 'reflect' is intended to indicate that the agent's final credence takes into account her IHC. On my favored way of understanding higher-order evidence, the agent's final credence would simply *match* her IHC. But I think that even plausible views on which higher-order evidence is taken as less powerful will depend on an independent reliability-assessment (see Christensen (2010, 2011). So I'll use 'reflect' to allow us to consider what that assessment should look like, without commitment to the exact way it would help determine the agent's final credence.

[5] Thanks to an editor of this volume for prompting me to clarify this point.

evidence on which the belief was based, and independence from the reasoning from that evidence to the initial belief).[6]

I think that the most promising approach will not exactly require independence from the relevant *first-order evidence*. For one thing, it is probably a mistake to divide items of evidence between the "first-order" and "higher-order" categories. Some items of evidence have both sorts of import, even with respect to the same belief. Consider a standard sort of example used to motivate conciliatory accounts of disagreement: I go to dinner with a friend and, after doing mental calculations based on the amount on the check, come to a confident conclusion about what our shares are. Then I learn that my friend disagrees. If we have long and equally-good track records in share-calculation, this seems to call for significant loss of confidence on my part. And it would intuitively beg the question for me to use my initial calculation as the basis for dismissing my friend's opinion as incorrect. But suppose we sought to prevent this question-begging by requiring my reliability-assessment to be independent of my "first-order evidence" (in this case, the figures on our check). And suppose further that I happen to know that my friend is particularly bad at calculations when the figure on the check has 7s in it, and that today's check contains three 7s. Here, it seems that I should take this information into account in my reliability assessment, and not lose so much confidence in my original answer. But this would be prevented if my reliability assessment had to be independent of the evidence of the figures on the check.[7]

This suggests that what needs to be "bracketed" by my reliability-assessment is not evidence E as such, but certain routes by which E might support P, and thereby support the claim that I was reliable on this occasion. Although it is often convenient to talk about "first-order evidence" vs. "higher-order evidence," we should—at least when we are being careful—distinguish instead between different *ways that a piece of evidence can bear* on the doxastic attitude it's rational for an agent to take toward a given proposition. Roughly, evidence bears on the rationality of an agent's attitude in a *higher-order* way insofar as it bears on that attitude indirectly, *via* bearing on how the agent should assess her reliability in assessing the direct bearing of her evidence on the relevant proposition.[8] So Alicia's information about being drugged bears on her rational credence in P in a higher-order way. In the restaurant-check example we just considered, the number on the check bears on my credence in my answer in both higher-order and first-order ways. I think that disagreement cases typically illustrate these dual roles that a single item of evidence can play: when I believe P on the basis of E, the disagreement of a friend who shares E serves both as first-order evidence for not-P (in the usual testimonial way), and as higher-order evidence that I've reasoned from E to P unreliably. This rough characterization is, no doubt, inadequate. But I think that the examples can at least serve to flesh out the intended idea, and that speaking in terms of different ways evidence can bear on an attitude, rather than speaking of different kinds of evidence, is a step in the right direction.[9]

---

[6] For a sampling of various formulations by defenders of such principles, see Elga (2007), Kornblith (2010), Christensen (2011), Vavova (2014), and Matheson (2015).

[7] Thanks to Adam Elga for helpful discussion of this sort of case.

[8] This differs from some common ways of characterizing higher-order evidence, which see it as directly bearing on the *rationality*, rather than the reliability, of the agent's assessment of the evidence's direct bearing on the relevant proposition. I suspect that our concern with rationality in these contexts derives from a more fundamental concern for accuracy; for more on this distinction, see Christensen (2014).

[9] Having made this distinction, I will often revert to informal talk of first-order and higher-order evidence, to avoid cumbersome formulations.

Suppose, then, that our Independence principle should focus on the train of reasoning from E to P, not on E itself. This fits with informal formulations which require the assessment to be "independent of the reasoning behind the initial belief." But this raises another question: what does it mean for, say, Alicia's reliability assessment to be independent of "the reasoning behind" her initial belief? Does it just mean that the assessment must not depend on the particular *episode of reasoning* in which Alicia engaged at the beginning of the story? Or does it mean that the reliability assessment that would be rational for Alicia should be independent of the general fact that A and B support P—that is, independent of the *first-order bearing* of A and B on P? The former interpretation would see the relevant reliability assessment in terms of the conditional credence that would be rational to have in P, given that Alicia was drugged, and that she'd reached a high initial credence in P, *and that A and B—which entail P—are very likely true.* That is, we would be allowing the reliability assessment that was rational for Alicia to be informed by A and B's actual first-order support for P. Since A and B entail P, that rational conditional credence would presumably be very high. But that's not the result the account is after; it really defeats the purpose behind requiring Independence. So I think that the latter interpretation is closer to what we need.[10] The reliability-assessment that would be rational for Alicia should be fully independent of A and B's first-order bearing on P. In general terms, then, we might amend our rough characterization as follows:

> **Independence, second sketch:** When an agent has formed an initial credence c in P on the basis of the first-order bearing of evidence E, and then gets some evidence that bears on the reliability of her reasoning from E to her credence c in P, her final credence in P should reflect the Independent Hypothetical Credence (IHC) it would be rational for her to have in P: that is, the rational credence in P *independent of E's first-order bearing on P*, but conditional on her having formed credence c in P on the basis of E, and on the reliability evidence the agent has about herself.[11]

This formulation avoids having to divide items of evidence into first-order and higher-order. It also allows for the possibility of agents using *facts about* their evidence or reasoning to inform their reliability-assessments. So, for example, in Alicia's case, if the drug did not affect cognition about people whose names began with K or L, our formulation would allow Alicia to take that into account and raise her assessment of her reliability. This fits well with a more general point that has been made in the literature: "independent of the reasoning behind the agent's initial belief" cannot mean that *facts about* this reasoning are precluded from informing the reliability-assessment. Agents should be allowed to take into account whether, e.g., they felt clear-headed while doing the reasoning, whether the reasoning began with a number involving 7s, whether it involved statistical analysis, etc. Independence principles require a distinction between depending on facts about my reasoning, and depending on that reasoning itself in a way that relies on the reasoning's cogency.[12]

---

[10] This distinction is developed in van Wietmarschen (2013), which argues that only the former interpretation can be motivated, and that seeing this shows that conciliatory accounts of disagreement apply only to the well-foundedness (doxastic justification) of beliefs, not to their evidential support (propositional justification). I resist these conclusions, and discuss the issue in greater depth, in Christensen (forthcoming). See also Smithies (2015) for a view on which higher-order evidence is relevant to doxastic, but not propositional, justification.

[11] I say "bearing on" rather than "support for" to cover cases where first-order evidence tells against the relevant proposition, or where it would make rational a mid-range credence.

[12] For more on this issue, see Christensen (2011), Arsenault & Irving (2012), Kelly (2013), Matheson (2015), Christensen (forthcoming).

This feature of our formulation will help handle a type of example due to Andrew Moon (2018). Moon develops a series of counterexamples to the sort of Independence principles that have been invoked in the disagreement literature—ones which require that one's assessment of the disagreer's reliability be "independent of the disputed belief" or "independent of the reasoning behind the disputed belief." Here is a representative example, adapted slightly:

> **Reliable Source**: Boris knows Cho to be a reliable source. Cho tells him two things:
> (P) Peggy is at the party.
> (Q) If Peggy is at the party, then Quinn is unreliable about Peggy's whereabouts.
> Boris rationally becomes confident in both P and Q. But then, Boris meets Quinn, about whom he has no other special information. Quinn says that Peggy was not at the party. Boris puts P and Q together, dismisses Quinn's testimony, and retains his belief that P.

Moon argues that Boris's reaction of dismissing Quinn's disagreement here is legitimate. But he argues that it relies crucially on Boris's using his beliefs that P and Q to support Quinn's unreliability, and thus dismiss Quinn's disagreement. And P is the very belief disputed by Quinn. So formulations of Independence which forbid this sort of dependence must be wrong. He proposes building in *exception clauses* to Independence principles which would permit reliability-assessments to be based on the disputed belief (or the reasoning behind it), as long as certain conditions obtain.

I think that there are reasons to worry about building in these exception clauses to our principle.[13] But it is worth noticing that if we understand Independence in the way we are examining, Reliable Source seems to be handled correctly. Boris's IHC should be independent of his evidence's first-order support for P, but will be based on the fact that he formed a high initial credence in P, and on the evidence he has that's relevant to his reliability. And it seems to me that the IHC will indeed be pretty high here, because of considerations involving *facts about* Boris's original reasoning: Boris formed a high credence in a proposition that was asserted by a source he knew to be reliable. The proposition was then denied by a source he knew nothing special about, and, in addition, the known-reliable source provided evidence that the other source was unreliable in cases where the proposition was true. It would seem that, *in cases fitting this general description*, the initial high credence is very likely to be accurate. But this reasoning does not depend on the claim that Peggy is at the party—it just relies on facts about the *types of evidence* Boris based his credence on, and on how these types of evidence interact.[14] If our formulation handles this sort of example correctly, perhaps we need not add potentially troublesome exception clauses to our Independence principle.

---

[13] Here is an example of why I think we should worry. One of the exception clauses permits the agent to rely on premises behind her initial belief to reach conclusions about the epistemic credentials of her friend's belief, as long as they concern the *non-truth-entailing* credentials of that belief. This is needed because Moon is sympathetic to the basic motivation behind the Independence principles in the literature. So he wants to disallow the sort of question-begging dismissal of disagreement that occurs in standard restaurant-check-style cases when I start out believing P, find out that my peer believes not-P, and reason that since P, her belief is *false* this time, and hence not warranted/known--so I needn't lose confidence in mine.

But consider a weather-forecasting peer disagreement case, where we build in that my peer and I each very occasionally become over- or under-confident in rain the next day. If I believe it'll rain tomorrow, and she, on the basis of the same meteorological evidence, thinks it won't, I should not be able to reason as follows: "Well, our data make rain highly probable, so she must have misinterpreted the data today, so I needn't worry about her disagreement". But the downgrading of my friend's belief's credentials in this case is not based on any claim entailing the truth or falsity of the claim that it will rain.

[14] I should note that Moon's description of Boris's reasoning to his conclusion about Quinn's reliability—which explicitly moves from "P" and "If P, then Quinn is unreliable about Peggy's whereabouts" to "Quinn is unreliable about

3. Independent of the first-order bearing of *what evidence* on P?

One of the ways which Alicia's case is particularly simple is that the first-order evidence bearing on P is cleanly isolated: as a first approximation, A and B bear on P, and nothing else does.[15] So in thinking about her reliability estimate, it seemed obvious that it should simply be independent of the first-order bearing of *A and B* on P. But if we are to find a general description of the requisite sort of independence, we will have to think about more problematic cases. Consider, for example:

> **Dinosaur Disagreement**: Dora and Esa disagree about the extinction of the dinosaurs: Dora believes that it was caused by volcanoes, while Esa thinks it was caused by a meteor impact. Let us suppose that each of them is rationally quite confident that the other is her (at least rough) equal in intelligence, education, diligence, honesty, acquaintance with the relevant literature, and the other factors that they take to bear on likely reliability—each reasonably thinks that the other would be just as likely to get the right answer to this sort of scientific question as she herself is.

Suppose we think that Dora should lose substantial confidence in her belief, given Esa's disagreement. We would like Dora's relevant Independent Hypothetical Credence in the volcano hypothesis—independent of her views on dinosaur extinction, but given what she has reason to believe about her own and Esa's reliability—to be relatively modest. But in this case, there's no simple way of isolating the first-order evidence E which bears on these hypotheses: clearly, all sorts of geological, biological, astronomical and archaeological evidence are relevant. And similar points will apply to lots of disagreement cases—e.g., philosophers' disagreements about the nature of consciousness, or economists' disagreements about the effects of minimum-wage laws. So simply specifying that Dora's IHC is independent of the first-order bearing of "evidence E" on the volcano hypothesis looks problematic.[16]

Moreover, even when there is some clearly-specifiable bit of relevant first-order evidence, there are sorts of cases where the proposed approach would not seem to fit. There are cases where the very same bit of evidence bears on the relevant proposition in more than one first-order way. In some such cases, the agent's higher-order evidence will threaten her ability to respond to some of these ways, but not others. So will the right reliability-assessment be independent of the first-order bearing of this evidence, or not? Consider the following example:

> **Two Diseases:** Freny is a doctor whose patient has symptoms that might indicate either of two diseases D1 or D2 (though he may have neither). Freny orders a blood test, and examines the results. They eliminate D1 simply and decisively, and also eliminate D2, but through a more complex type of statistical reasoning. Freny forms very low credences in

---

Peggy's whereabouts"—does seem quite natural, intuitively. But I think that this should not persuade us that Boris may really rationally rely on his belief that P to dismiss worries about his reliability in reasoning about P. See Christensen (forthcoming) for detailed discussion of this issue.

[15] This isn't quite right. Alicia's very high credence in A and B is itself based on the reliable source's testimony. But I believe that this temporary simplification is harmless.

[16] See Christensen (2011, 18ff.) for an expression of this worry, whose importance was emphasized to me by Jennifer Lackey.

both D1 and D2. Then she's informed that she's been dosed with a drug that doesn't affect the sort of simple thinking required to eliminate D1 on the basis of the blood test, but that does render more complex statistical thinking very unreliable. It seems that Freny should end up with high confidence that her patient is free of D1, but not with high confidence that her patient is free of D2.

It's hard to see how the difference between the way Freny's evidence bears on D1 and the way it bears on D2 would be captured by separating this evidence into two chunks: it's the same blood test result that grounds both chains of reasoning. But how else could the independent hypothetical credence that would be rational for her reflect one avenue of support but not the other?

A clue to one approach can be found in thinking back to Alicia's case. There, we saw that Alicia's IHC in P was .7. This was not, of course, the credence directly supported by her first-order evidence (being told that A and B) alone; that was the point of imposing an Independence requirement. But Alicia's .7 hypothetical credence was also not *unaffected* by Alicia's having that evidence. It was conditional on her having reached a very high initial credence in P—an initial credence she would not have reached had she not had A and B to go on. And it is only the fact that her initial credence was so high—combined with the information about the drug—that made her IHC .7, rather than some much lower value. Moreover, if the drug had been much weaker, Alicia's IHC would have been correspondingly closer to .99. If the drug had had only a negligible expected effect on Alicia's thinking, her IHC would have been extremely close to the .99 level made rational by the first-order bearing of her evidence—even though the IHC was independent of A and B's first-order support for P. So: when a bit of first-order evidence is part of the evidence whose first-order bearing on the relevant proposition is "bracketed" by the IHC, *the fact that the agent has that evidence* may still affect the agent's IHC, and thus affect the agent's rational final credence.[17]

This suggests that one approach to the Two Diseases case would involve taking Freny's IHC to be independent of the first-order bearing of the blood-test results *entirely*. The idea is that the fact that Freny reached certain credences on the basis of the drug test might still allow the blood-test data to inform her IHC in an intuitively reasonable way.

So let us ask, first, what credence Freny should have that her patient is free of D2, independent of the blood test's first-order bearing on D2, but conditional on Freny having formed low credence in D2 on the basis of *statistical* reasoning from the blood test, and on Freny being a generally reliable doctor who was drugged in a way likely to make that sort of reasoning unreliable. Here, it seems that Freny's IHC that her patient is free of D2 will not be high, since her elimination of D2 is likely to have been based on faulty reasoning. This seems like the appropriate result.

Next, let's ask what credence Freny should have that her patient is free of D1, independent of the blood test's first-order bearing on D1, but conditional on Freny having formed low credence in D1 on the basis of *simple* reasoning from the blood test, and on Freny being a generally reliable doctor who was *not* drugged in a way likely to make *that* sort of reasoning unreliable. Here, it seems that Freny's IHC that her patient is free of D1 will be high, since her elimination of D1 is unlikely to have been based on a mistake. This is also the result we want.

So despite our stipulation that Freny's IHC be independent of the first-order bearing of the blood-test evidence, the blood-test evidence did affect the IHC it was rational for her to have in D1. The fact that she formed her initial low credence in D1 on the basis of that evidence, combined with the fact that her higher-order evidence indicated that she was in a good position to assess the

---

[17] Recall that we are working with two important simplifying assumptions: that the agent reached an initial credence on the basis of her first-order evidence before getting her higher-order evidence, and that that initial credence was rational. We will turn to relaxing these assumptions below.

bearing of that blood-test evidence on D1, led to an IHC that reflected, indirectly, the blood-test's elimination of D1.

On the other hand, the blood-test evidence did not similarly affect Freny's IHC for D2. While Freny did in fact appreciate its significance for D2 correctly, her higher-order evidence did not allow her to trust that initial low credence in D2. So perhaps we can after all handle cases where a single piece of evidence bears on the agent's beliefs in some first-order ways that are threatened by the agent's higher-order evidence, and in other first-order ways that are not: we can take the relevant IHC to be independent of *all* the first-order bearing of that evidence on the relevant propositions.

If this technique allows us to recover the effect that the evidence legitimately has when it bears on matters in ways that aren't threatened by the agent's higher-order evidence, the same technique might be extended to offer a neat solution to the problem posed by Dinosaur Disagreement. There, it looked like we'd need to find a way of delimiting exactly what first-order "evidence E" was relevant to Dora's belief in the volcano hypothesis. But perhaps we don't need to do that after all. Instead, we might make the relevant IHC independent of the first-order bearing of *all* evidence on the volcano hypothesis. Insofar as Dora's higher-order evidence does not cast doubt on her ability to interpret certain parts of that evidence, her earlier appreciation of that evidence will affect the IHC it's rational for her to have, in the same way as the blood test's elimination of D1 ended up affecting Freny's IHC in D1.

This suggests a further refinement of our rough formulation of Independence:

> **Independence, third sketch:** When an agent has formed an initial credence c in P on the basis of the first-order bearing of her evidence, and then gets some evidence that bears on the reliability of her reasoning from that evidence to her credence c in P, her final credence in P should reflect the Independent Hypothetical Credence (IHC) it would be rational for her to have in P: that is, the rational credence in P *independent of her evidence's first-order bearing on P*, but conditional on her having formed credence c in P on the basis of *the first-order bearing of her evidence*, and on the reliability evidence the agent has about herself.

This formulation has one other advantage worth noting. It allows us to accommodate a kind of case that could otherwise seem problematic: the sort of case that arguably does not involve any first-order evidence for P. Suppose, for example, that we vary Logic on Drugs so that Alicia does not learn A or B from a reliable source. Instead, she just decides to consider the material conditional "if A and B, then P." Alicia, ever the good logician, recognizes this as a logical truth and becomes highly confident of it. Then she's told she's been drugged in a way that makes people like her misidentify truth-functionally complex claims about birthdays as logically true about 30% of the time. Alicia reduces her credence in the claim to around .7.

To get this result, we should see Alicia as having a rational IHC in the claim, conditional on her having reached a high credence in the claim, and on her having been drugged in a specific way, of around .7. And one might worry that since the claim is a logical truth, the rational IHC will have to be 1, no matter what it's conditional on.

But it seems to me that we need not say this. We might see logical truths as supported in a first-order way by *any* set of premises (including, of course, the empty set). After all, the way that A and B support P in the original version of our example is via the logical relations among the claims: the logical relations guarantee that P is true if A and B are. But these are the same logical relations that guarantee the truth of the material conditional whether or not A and B are true. So it seems natural that if we make Alicia's rational IHC in our original case independent of the first-order support A and B give to P, via those logical relations, we should consider Alicia's rational IHC in the

variant case to be independent of the bearing of these logical relations on the truth of the material conditional, no matter what Alicia's first-order evidence may be.[18]

4. What if there's "no higher-order evidence"?

So far, we've been thinking of cases where agents get clear evidence that some part of their thinking has been compromised. It might seem at first as if this is some fairly limited class of special cases that comprise the domain of theories of higher-order evidence. And the literature on higher-order evidence and disagreement often reads as if it's a discussion of somewhat anomalous cases.

But a little reflection should cast doubt on that thought. Consider some variations on Logic on Drugs. We might begin with a case where Alicia gets *excellent* evidence that she's been given a drug that *always* makes *everyone* reach wrong conclusions about birthdays. We can then construct one kind of spectrum by reducing the strength of Alicia's reasons for trusting the person who informs her about the drugs, until this evidence is extremely weak. Or we can consider cases where Alicia gets excellent evidence that she's been given a drug which always causes those sensitive to it to make mistakes, but which only affects 30% of people it's given to. Or one which only affects 1%, or even .01%. Clearly, in each of these spectra, the last cases should be treated much like cases where Alicia gets no special information about being drugged. So cases involving clear evidence of impairment seem to shade smoothly into those involving no such evidence.

We can also construct a spectrum of variant cases where Alicia's evidence concerns a drug that affects everyone, but which induces 30% inaccuracy, or 1% inaccuracy or .01% inaccuracy. This sort of spectrum can be continued to include cases where Alicia gets evidence that her reliability has been *enhanced* rather than compromised. (A realistic example of such a spectrum might involve beverages ranging from cognac to coffee.) So a general account of higher-order evidence should handle cases ranging from impairment to enhancement, presumably in a smooth way.

It's also worth noting that, insofar as Alicia is like an ordinary person, she will have some evidence relevant to her reliability, even apart from any drug information. This might come from how clear or fuzzy she felt in thinking about the birthday problem. It might come from her track record on very similar problems—or, less tellingly, from her vague impressions of her track record on less-similar problems. And so on. So if Alicia is an ordinary person, she likely will have some higher-order evidence bearing on her likely reliability. And this does not seem peculiar to logic problems—it applies to a vast range of cases where people draw conclusions from evidence.

All of these considerations together suggest that our account of higher-order evidence should apply to most, if not all, of the beliefs we form on the basis of first-order evidence. If that's right, then in a great many cases, our beliefs should be required to reflect our rational IHCs.[19]

However, it's not obvious how the relevant rational IHCs would be arrived at in cases where reliability evidence is meager. And it is hard to say much about this without presupposing one or another general account of epistemic rationality. But having flagged the problem, I'll sketch one approach that strikes me as attractive (though it no doubt incorporates assumptions that some would reject).

---

[18] This makes especially clear that our IHCs will not be conditional probabilities. But it seems clear that taking higher-order evidence seriously will lead to non-probabilistic credences in general. See Christensen (2007).

[19] At this point, I'm still restricting our discussion to cases where an agent forms a belief on the basis of first-order considerations, and asking how she should take higher-order considerations into account. Thinking about ordinary cases makes salient how severe a limitation this is. The next section takes on this problem.

Let us begin by considering the ordinary range of beliefs that ordinary adults typically have in ordinary matters—cases where agents aren't parties to disagreement, or likely victims of cognition-degrading factors such as drugs, biases, or fatigue. It does seem reasonable—and I think it will probably be uncontroversial—that these people are usually rational in trusting their own cognition. One way of putting this, very roughly, is to say that, in most cases, agents would be rational in thinking that P is likely to be true (false), on the supposition that they believe that P (~P). Less roughly, agents are typically rational to have high (low) credence in P being true, conditional on their having a high (low) credence in P.

Next, it seems that this sort of conditional credence would typically be rational, even independent of the agent's particular attitude toward P. Someone may believe P, and also may be rational to be confident that P is likely true on the condition that she believes it, and the rationality of this conditional credence need not be just based on her confidence that P is true anyway. So, for example, if we describe the belief about P in general terms ("a belief about which of two letters comes first in the alphabet", "a belief about what country a well-known city is in", "a belief about whether a certain kind of animal is a mammal"), agents will typically be rational to be confident that P is true on the condition that they believe that P.[20] If this is right, then even when we consider the ordinary beliefs of reasonable people—beliefs that are not subject to the sort of higher-order evidence that the literature has highlighted—we should typically see agents as rational in having Independent Hypothetical Credences that cohere with those ordinary beliefs.

There are, of course, deep and difficult questions about how this sort of self-confidence is ultimately grounded. One way of raising such questions begins by considering the theoretical limit case, where an agent has no higher-order evidence at all to work with. It's quite possible that such cases never occur in adult human thinkers, but perhaps we can ask the theoretical question of what reliability-assessment would be rational, completely independent of an agent's higher-order evidence. It seems plausible that when we ask ourselves this question, the answer is that agents have some default entitlement to trust their own thinking. This is a familiar idea, made plausible in part by the apparent impossibility of rationally basing beliefs on deliverances of sources whose reliability one cannot rationally trust. On this sort of picture, an agent with no higher-order evidence at all would be rational to have some (defeasible, of course) confidence in the accuracy of her beliefs.[21]

It's worth noting that this picture fits well with one kind of line in the disagreement literature: Some have worried that Independence principles will lead to skepticism in disagreements with, e.g., paranoid conspiracy theorists who reject vast proportions of our beliefs. The worry is that if we put aside *all* the disputed issues, we'll be left with no independent reason to think we're more reliable than the conspiracy theorist. In response, it has been argued that conciliation is required only when one has strong independent reason to believe the other person to be reliable, and that putting aside most of our beliefs, we're left without strong reason think this.[22] On the current suggestion, it's natural to think that when there's only very slim dispute-independent evidence to go on, it will be

[20] A more precise version would talk about credences rather than beliefs; this would require talking about cases where the agent has some middling credence in P. But the point here is just the intuitive one: that for most of us, and for many of the ordinary topics we have opinions about, we have reason for confidence in our opinions that is independent of those particular opinions themselves.

[21] See Sliwa and Horowitz (2015) for a similar suggestion in response to essentially the same problem for their account of higher-order evidence. See Foley (2001) for extended discussion of the general issue in epistemology. As Jonathan Vogel pointed out to me, the envisioned position is at least analogous to holding that rationality involves having certain priors for propositions about our reliability, as well as about other matters.

[22] See, e.g., Elga (2007), Christensen (2011), and Vavova (2014); related issues are discussed in Worsnip (2014) and Pittard (2017).

insufficiently robust to significantly undermine our default self-trust. So our default self-trust provides a graded mechanism whereby undermining the justification of our beliefs occurs only to the extent that we have substantial independent reason to trust those with whom we disagree.

But perhaps we don't need to settle this in order to say that there are IHCs that are rational for ordinary people who have ordinary beliefs. Of course, these will be highly sensitive to specific reliability information of the sort featured in standard discussions of higher-order evidence. And they clearly may be pushed in either direction by specific higher-order evidence, making room for evidence of enhanced reliability as well as the sorts of diminished reliability the literature has concentrated on.

5. What if the agent has no "initial credence"?

We've been simplifying the discussion by supposing that our agents have formed rational credences on the basis of their first-order evidence before they get the higher-order evidence. But surely this is not how matters typically work out. One typically forms opinions on controversial issues by looking at the first-order evidence while already aware of the fact that others have varying opinions. One may learn about implicit bias before looking at a stack of CVs, or learn about possible hypoxia before wondering whether one has enough fuel to reach Sitka.

In thinking about this sort of case, we should pay attention to the distinction between propositional and doxastic rationality. Here, I'd like to concentrate on the propositional notion: When we think about agents who get higher-order evidence before getting the first-order evidence, what credence is propositionally rational for them? And I will be supposing that what credence is propositionally rational for agents depends not only on their first-order evidence, but on their higher-order evidence as well.

Thinking about cases where the agent does not have an initial credence based only on first-order considerations will clearly complicate our discussion considerably. Our Independence principles so far have simply assumed that agents had already formed rational credences on their first-order evidence, and it was those credences that were the subject of the independent reliability-assessments. That clearly can't work in cases where the agent has not formed any initial credence. So what could our reliability-assessments possibly apply to?

Without delving into different possible approaches to this question, I'll assume for present purposes a particular answer to it, in order to lay bare the issues that come up for formulating Independence. The basic idea is this: since we're asking what final credence is rational given the agent's total evidence, we can apply the agent's higher-order-evidence-based reliability-assessment to the credence *that would be rational for her on the basis of the first-order bearing of her evidence*. In effect, we are imagining that the agent reacts rationally to her first-order evidence, and takes that reaction as the subject of the reliability-assessment. This would of course mesh nicely with our verdicts on the cases considered above, where we stipulated that the agent had already formed the credence that was rational given her first-order evidence. It also fits with the natural idea that an agent who gets higher-order evidence before first-order evidence should end up with the same credence as one who gets the same batches of evidence, but in the opposite order. [23]

Moving to applying reliability assessments to credences the agent hasn't actually formed, though, introduces another complication. When we were concentrating on agents who had already formed an initial credence in the relevant proposition P, the agent's IHC was conditional on the fact that she herself had formed that initial credence. But in the present framework, the agent's evidence presumably includes her memory of not having formed any such credence. This does not necessarily

---

[23] See Christensen (2016) for extended discussion of this idea.

mean that she would be rational to be absolutely certain that she did not form an initial credence—perhaps she is not. But she might well be rational to think that if she had formed such a credence, then given that she can't remember any such thing, there must be something drastically wrong with her. This would obviously suggest that she was cognitively impaired in some way, and thus affect her IHC in ways that we want to avoid.

Here is what seems to me to be the most promising way of avoiding this problem: Instead of considering an IHC conditional on the agent *herself* having formed the initial credence, we might consider an IHC that's focused on a *relevantly similar* hypothetical agent—that is, one who is similar to the agent with respect to those higher-order factors that affect expected reliability. (In this way, the relevant hypothetical credence is similar to what would be rational for a third party to form, upon learning that an agent of the relevant sort had formed the initial credence in question.) Putting these ideas together, we get something along the following lines:

> **Independence, final sketch:** Let c be the credence in P *that would be rational* given the first-order bearing of the agent's evidence on P. Then the credence in P that would be rational for the agent, given *all* her evidence, should reflect the Independent Hypothetical Credence (IHC) it would be rational for her to have in P: that is, the rational credence in P independent of her evidence's first-order bearing on P, but conditional on a *relevantly similar* agent adopting credence c in P on the basis of the first-order bearing of that agent's evidence.

There are, no doubt, further problems with this sketch—so it's "final" only in representing the approach that looks most promising to me today. But before closing, in the last two sections I'd like to discuss one apparent difficulty that this formulation poses, and another difficulty that becomes apparent when we move beyond the simplified examples we've been looking at so far.

6. What if there are multiple first-order routes from E to P?

Agents who have actually formed initial credences are, of course, aware of more than the fact that they formed a certain initial credence. They may well be aware, for instance, of the way in which they moved from their first-order evidence to their initial credence in P. For example, in a case like Logic on Drugs, Alicia may well remember which logical rules she actually used initially, in inferring P from A and B. In our original case, this information was not particularly relevant, since the drug we were imagining affected all complex truth-functional reasoning about birthdays. But designing our drugs a bit more narrowly can give us cases where this sort of information can be relevant. For example, suppose that Alicia knew that the logic-disrupting drug she'd been given only affected reasoning done by (attempted) Modus Ponens. In such a case, the confidence it would be rational for her to end up with would depend on whether she had reason to think she'd relied on MP in forming her high credence in P. This might not raise any difficulties in Alicia's case, but we should also consider cases where the agent does not form their initial credence before getting the higher-order evidence. So consider the following case:

> **Logic on MP-disrupting Drugs:** Gabi is informed that they have been slipped a drug that degrades people's truth-functional reasoning about birthdays. But this drug only degrades Modus Ponens reasoning: those who attempt to reason by MP make mistakes in applying the rule 30% of the time, while feeling perfectly clear-headed. The drug does not affect other

reasoning at all. After learning about the drug, Gabi is told A and B, as Alicia was, and then asked whether P.

What should Gabi's IHC be? If Gabi had inferred P from A and B before getting the information about the drug, they would have information about how they had done that—information relevant to the reliability of their inference (i.e., whether they had used MP in deriving P from A and B). But in the present case, Gabi never performed that initial inference. So in thinking about what credence in P is propositionally rational for Gabi, how should the information about the drug be brought to bear?

We should first note that the credence c that's rational on Gabi's first-order evidence is very high, since A and B entail P. So our question then becomes: how likely is P to be true, given that an agent relevantly like Gabi reached a very high credence in P on the basis of first-order support from their evidence? The answer to this question obviously depends on how likely that agent would have been to reason via MP. If the agent was highly likely to have reasoned via MP, it will be close to 30% likely that they made a mistake, so the IHC would be significantly lower than c. On the other hand, if the agent would be unlikely to use MP, the probability of mistake would be much lower, and the rational IHC much closer to c.

The next thing to notice is that the reliability evidence Gabi has does not only make it likely that Gabi has been drugged in a certain way. It also includes the fact that *Gabi was told* they had been drugged in that way. If Gabi is like most of us, the fact that they have good reason to think they would be likely to mess up in MP-based reasoning would make them much more likely to try reason about the problem without using MP. If it would be reasonable, given Gabi's evidence about this, to expect Gabi—or a relevantly similar agent—to be able to succeed in this task, then the IHC that would be rational on Gabi's evidence would presumably be quite high.

On the other hand, suppose that Gabi's evidence suggests that they are so drawn to reasoning by MP that they would end up slipping and reasoning by attempting MP, without realizing it, even if they tried not to. Then the same would go for the relevantly similar agent. And of course the rational IHC for Gabi in this situation would be significantly lower. So it turns out that the IHC that would be rational for Gabi would depend on how likely it was (on their evidence) that they would end up reasoning by attempted MP.

This might seem problematic at first. After all, one might think, if we're asking what credence is propositionally rational for an agent, we should answer this by assuming that they react to their evidence rationally. And given Gabi's evidence about having ingested the drug, the rational reaction would include avoiding MP. So despite what evidence Gabi has about their own propensity to use MP, their IHC in P would be very high. But intuitively, in cases where Gabi has good reason to believe they would reason unreliably, high credence in P would not be propositionally rational.

But I think this worry would be misplaced. At bottom, it's not really different from the general thought that higher-order evidence is always irrelevant to the question of what credence is propositionally rational, since a perfectly rational agent would react perfectly to her first-order evidence. Insofar as we acknowledge that agents are rationally required to take higher-order evidence seriously—in the sense that what credences are propositionally rational for them are sensitive to information they have about their own expected reliability—it makes sense that the credence in P that's rational for Gabi will depend on how likely Gabi's higher-order evidence makes it that they will reason unreliably. Gabi's situation is not really different in kind from the situations of agents who have evidence that they might be hypoxic, or biased. In all of these cases, the credence that's propositionally rational for the agent depends on the agent's evidence about their own reliability.

So it seems to me that while moving to our third sketch of Independence does introduce complexities in thinking about cases involving multiple paths from E to P, the complexities simply reflect important complexities of the epistemic situations in question.

7. What if the higher-order evidence is indiscriminate in its target?

There is one more way in which the examples we've been looking at were chosen to simplify our discussion of Independence principles. In each case, the higher-order evidence targeted a fairly narrow part of the agent's reasoning. This is particularly clear in our artificial cases involving designer drugs, but also applies in other cases. Sexism may distort my CV-assessments while leaving large parts of my cognition unaffected. Fatigue or hypoxia may degrade complex thinking, while leaving simpler thinking relatively unscathed. Disagreement about a particular topic may indicate that I've made a mistake in thinking about that topic, without necessarily indicating a more widespread cognitive problem.

But it's clear that not all cases will fit into this tidy mold. Sometimes, higher-order evidence may provide reason for the agent to doubt wide swaths of her thinking—including, in some cases, her thinking about how to accommodate higher-order evidence. This sort of case introduces a new dimension of difficulty in theorizing about higher-order evidence.[24]

For a simple example of this sort, we might think about more powerful drugs. Suppose that Hui is part of a research team studying a new reason-disrupting drug. The team has seen subjects like Alicia and Gabi reach reasonable conclusions about P—conclusions that reflect their higher-order reasons for doubting their ability to reason truth-functionally about birthdays while drugged. So they decide to kick the challenge up a notch. They design a drug that not only severely compromises ordinary truth-functional reasoning about birthdays, but also degrades the sort of thinking required to take higher-order evidence about drugs into account—in particular, drugged agents do not form the IHC's supported by their evidence. Hui volunteers to be a subject for the initial tests of the drug. She walks into the experiment room, and an associate brings her a pill and some water. 20 minutes later, she is given some information, and asked to think about what follows. Suppose she is told A and B from our original story, and wonders whether P. It sure *seems* to her as if P must be true if A and B are—but….

Well, what credence in P is rational for Hui in this situation? In thinking intuitively about this case, we might start by asking how a fully rational agent would react to Hui's predicament. One might think that since Hui has good evidence that she can't rationally take higher-order evidence into account, she should just go with her first impression, and be highly confident in P. But we might imagine that she's been through several trials with this drug, and knows that this strategy very frequently leads to high confidence in false claims—after all, the drug degrades first-order thinking severely, so she can't trust that thinking any more than she can trust her thinking about the implications of her higher-order evidence. So maybe she should take all of this into account, and ignore not only her higher-order evidence, but also ignore A and B. But would not that be simply a different instance of using higher-order evidence to compensate for the drug's expected effects? (In fact, even going with her initial impression that P follows from A and B, on the grounds that she can't trust herself to reliably form beliefs on the basis of the drug information, would also be an instance of taking the drug information into account in forming her belief.) At this point, I, at any rate, have no clear intuitive idea what maximally-rational Hui would believe, or what result the right account of higher-order evidence *should* give in this case.

---

[24] Sliwa and Horowitz (2015, §5.1) pose this problem, and make points related to some of the points made below.

Does Independence help us out here, giving us a verdict in a case where intuitions are confused? I don't think so. Independence—at least as it's sketched above—invokes *rational* Independent Hypothetical Credences. In this case, we would have to ask how confident it's rational for Hui to be in P, independent of P's first-order support, but contingent on an agent relevantly similar to Hui reaching high credence in P on the basis of first-order bearing of that agent's evidence. But this assessment is exactly the sort that's targeted by Hui's higher-order evidence. And so it's not clear what IHC would be rational for Hui in this case.

Should we *expect* the correct account of Independence to deliver a clear verdict about what credence is rational for Hui's? I think it's not at all obvious that we should. Independence feeds into an account of higher-order evidence which puts constraints on rational credences: it relates rational "all things considered" credences to certain rational hypothetical conditional credences. But this constraint need not tell us the whole story about what either of these two credences should be. Insofar as Hui's case is one where it's quite unclear intuitively what the most rational overall credence, or the rational IHC, would be, it is less worrisome that Independence yields no clear verdict. We might consider an extreme version of Hui's case, in which a drug is so powerful that it completely messes up people's thinking about everything. It seems that Hui's team could observe the wacky beliefs their unwitting experimental subjects form under the drug's influence—but it's not at all clear that this leaves Hui with some epistemically rational reaction to the information that *she's* just ingested that drug. A similar problem crops up in examples posing the anti-expertise paradox, where agents get excellent evidence that they'll believe P just in case P is false: no doxastic position seems rationally stable.[25] It would seem unreasonable to expect Independence to provide a cure for this sort of epistemic malady.

Nevertheless, I think that cases involving indiscriminate higher-order evidence do pose problems for formulating Independence. Suppose we decide that we need not worry about certain paradoxical cases, or cases involving near-global self-doubt. Still, cases of global self-doubt seem to be at one end of a spectrum, with cases of narrowly focused self-doubt at the other end. One might hope that a more fully developed account of higher-order evidence could give us some insight into the rationality of beliefs based on what we might think of as *somewhat indiscriminate* higher-order evidence. And some such examples are much less artificial than ones involving super-drugs, or the ones that standardly figure in discussions of the anti-expertise paradox.

For example, consider Isaac's attitude toward the proposition that the President of his country is a habitual liar. When Isaac considers the matter directly, it seems quite clear to him that his President is a habitual liar. But Isaac also notices that he's highly disgusted by many aspects of the President's conduct and character. He sees the President's policies as cruel and vindictive; he sees the President as surrounding himself with corrupt, self-serving advisors; and he sees the President as playing to his compatriots' worst selves. Overall, he thinks of the President as base and immoral, and is deeply embarrassed that his country is governed by such a man: even hearing the sound of the president's voice on the radio, or seeing his picture in the newspaper, turns Isaac's stomach. So the topic of the President is emotionally fraught for Isaac, and he's aware that his intense loathing for the President may compromise his ability to assess the President's truthfulness accurately.

It seems clear that this information is relevant to how confident Isaac may rationally be that the President is indeed a habitual liar. On the approach we've been developing, Isaac's rational credence depends on what IHC would be rational for him, given his evidence about his likely cognitive impairment. But should Isaac expect himself to be able to assess his likely reliability *on this matter* in a cool, accurate way? I don't think so. We may suppose that Isaac is sophisticated enough

---

[25] See Conee (1987), Sorensen (1987), Christensen (2010).

to see that his opinion on the President's mendacity and his opinion on his own reliability in assessing the President's mendacity are tightly linked. Insofar as his emotional investment in thinking the President a liar is likely to affect one, it seems quite likely to affect the other. So the rational IHC for Isaac would have to take into account his evidence that he's likely to *count himself more reliable* than his independent reliability-relevant evidence really indicates.

A parallel point applies to disagreement evidence. Suppose Isaac also knows that a significant number of his compatriots think that the President is not a habitual liar. How strong evidence does their disagreement provide for the claim that Isaac's direct thinking on the matter is unreliable? Well, as usual, that depends on how reliable Isaac should take those who disagree to be—setting aside Isaac's opinions on the President's truthfulness, and on related matters on which he disagrees with them. So, for example, Isaac might want to consider whether he has good independent reason to think that his opponents are being irrationally manipulated by rhetoric, or whether he has good independent reason to suspect that their views are formed in response to racism, or xenophobia.

But of course there's a problem here. Since Isaac is aware that his opponents' expected reliability is a threat to his own convictions, his emotional investment in those convictions would seem likely to warp his reliability-assessments of them (which of course feed directly into how reliable Isaac should expect his own contrary opinion to be).

Insofar as Isaac realizes all of this, it seems that that the rational IHC for him would be lower than it would be if Isaac did not have evidence that his emotions were likely to inflate his self-assessment. So this is not a case where we have no intuitive idea of how the agent's higher-order evidence should affect his credence. For that reason, it seems plausible that a fully satisfactory account of Independence should explain how an agent's evidence can bear *in a higher-order way* not only on what credences in ordinary propositions are rational for the agent, but on what IHCs are rational for the agent.

It's not clear to me how this will work out. When we ask how much Isaac's evidence about his emotions should lower his IHC, it seems that this question should depend on how reliable Isaac should expect himself to be at arriving at the rational IHC. This might naturally seem to involve some sort of meta-IHC. But it's not clear to me how this should be formulated. And it's not clear where, or if, Isaac's reasons for self-doubt will stop ramifying.[26]

What should we make of this problem?

We should not, of course, expect an Independence principle to give us an entire epistemology. We should not expect it to yield clear verdicts in every case, if there are cases in which there are no clearly correct answers. But it's not unreasonable to hope that our account of Independence will help us understand how higher-order evidence affects rational IHCs—that it will give us some insight into how ramified self-doubt affects rational credences, when those credences do seem to exist. I'm at this point unsure of how difficult this work will be. But at a minimum, there is more work to be done.

8. Conclusion

Independence principles seem to be required in order to cope with a peculiarity that characterizes higher-order evidence: roughly, this sort of evidence targets the reliability of an agent's thinking; but in assessing the import of this evidence, the agent must do so from within an epistemic perspective that's constituted by her own thinking. She must act both as judger and as judged.

---

[26] See Lasonen-Aarnio (2014) for an extended development of a related sort of worry, and Schechter (2013) for a possible line of response.

The awkwardness that this involves comes out, in various ways, in the difficulties we've seen in formulating Independence principles. While I think that there are promising avenues for meeting some of these difficulties—as embodied in the Independence-sketches above—I would not claim to have a clear, clean, and precise way of meeting them all.

One reaction one might have to the difficulties is to give up on the project—say, by holding that higher-order evidence simply does not bear on rational belief. But the gain in theoretical simplicity would, I think, be more than offset by the loss of explanatory power. When we consider the high-flying pilot, or the sleepy medical resident, or the person likely to be affected by sexist bias, it seems clearly irrational for them to maintain high confidence in their original judgments.

Like it or not, we're often cast in the role of judging our own epistemic performance. Acting as judge in one's own case is, of course, famously problematic. But since epistemic agents cannot recuse themselves from this predicament, we epistemologists should do our best to understand how it is most rationally negotiated.

David Christensen

References

Arsenault, M. and Z. C. Irving, (2012), "Aha! Trick Questions, Independence, and the Epistemology of Disagreement," *Thought* 1 (3): 185–194.

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

---. (2010), "Higher-Order Evidence, *Philosophy and Phenomenological Research*, 81 (1): 185–215.

---. (2011), "Disagreement, Question-Begging and Epistemic Self-Criticism," *Philosophers Imprint*, 11(6): 1–22.

---. (2014), "Conciliation, Uniqueness and Rational Toxicity," *Noûs* 50, 3: 584-603.

---. (2016), "Disagreement, Drugs, etc.: from Accuracy to Akrasia," *Episteme*: 397-422.

---. (forthcoming), "On Acting as Judge in One's Own (Epistemic) Case (Marc Sanders Lecture)," *Proceedings and Addresses of the American Philosophical Association.*

Conee, E. (1987), "Evident, but Rationally Unacceptable," Australasian Journal of Philosophy 65: 316 – 326.

Elga, A. (2007), "Reflection and Disagreement," *Noûs*, 41: 478–502.

Foley, R. (2001), *Intellectual Trust in Oneself and Others*, Cambridge University Press.

Kelly, T. (2013), Disagreement and the Burdens of Judgment," in D. Christensen and J. Lackey (eds.), *The Epistemology of Disagreement: New Essays*, Oxford University Press.

Kornblith, H. (2010), "Belief in the Face of Controversy," in Richard Feldman and Ted Warfield (eds.), *Disagreement*, Oxford University Press.

Lasonen-Aarnio, M. (2014), "Higher-Order Evidence and the Limits of Defeat," *Philosophy and Phenomenological Research* 88.2: 314-345.

Matheson, J. (2015), *The Epistemic Significance of Disagreement*, London: Palgrave Macmillan.

Moon, A. (2018), "Disagreement and New Ways to Remain Steadfast in the Face of Disagreement," *Episteme* 15, 1: 65-79.

Pittard, J. (2017), "Disagreement, Reliability and Resilience," *Synthese* 194: 4389-4409.

Schechter, J. (2013), "Rational Self-Doubt and the Failure of Closure," *Philosophical Studies* 163 (2): 428–452.

Sliwa, P. and S. Horowitz (2015), "Respecting *all* the evidence," *Philosophical Studies* 172 (11): 2835-2858.

Smithies, D. (2015), "Ideal Rationality and Logical Omniscience," *Synthese* 192.9: 2769-2793.

Sorensen, R. A. (1987), "Anti-Expertise, Instability, and Rational Choice," Australasian Journal of Philosophy 65: 301 – 315.

van Wietmarschen, H. (2013), "Peer Disagreement, Evidence, and Well-Groundedness," *Philosophical Review* 122 (3): 395–425.

Vavova, K. (2014), "Moral Disagreement and Moral Skepticism," *Philosophical Perspectives* 28 (1): 302–333.

Worsnip, A. (2014), "Disagreement about Disagreement? What Disagreement about Disagreement?" *Philosophers Imprint* 14 (18): 1–20.