Higher-Order Uncertainty

Kevin Dorst

kmdorst@mit.edu

July 2018

Abstract

You have *higher-order uncertainty* iff you are uncertain of what opinions you should have. I defend three claims about it. First, the higher-order evidence debate can be helpfully reframed in terms of higher-order uncertainty. The central question becomes how your first- and higher-order opinions should relate—a precise question that can be embedded within a general, tractable framework. Second, this question is nontrivial. Rational higher-order uncertainty is pervasive, and lies at the foundations of the epistemology of disagreement. Third, the answer is not obvious. The Enkratic Intuition—that your first-order opinions must "line up" with your higher-order opinions—is incorrect; epistemic *a*krasia can be rational. If all this is right, then it leaves us without answers—but with a clear picture of the question, and a fruitful strategy for pursuing it.

Keywords: Higher-order evidence; disagreement; epistemic akrasia; probabilistic epistemic logic.

Here is one of my main claims:

> *Thesis:* Epistemic akrasia can be rational.

(Don't worry, just yet, about what it means.) I am confident of *Thesis*, for I have a variety of arguments that I take to be good evidence for it.

But—now that I think about it—*whenever* I sit down to write a paper, I'm confident of that paper's thesis. In fact, that confidence usually has a similar basis: I have a variety of arguments that I take to be good evidence for it. And yet I've later found—all too often—that my arguments weren't so good after all; that I've been *over*confident in my past theses.[1] Having meditated on these facts, I'm still confident of *Thesis*, for I still think that I have good arguments for it. However, here's another proposition that I now consider possible:

---

[1] 'Overconfident' here—as in natural language—means being more confident than you *should* be; not to having some confidence in something that's false. If you're 50-50 that this fair coin that I'm about to toss will land heads, then you're not overconfident—even if, in fact, it will land tails.

Doubt: I should not be confident of *Thesis.*

I'm not confident of *Doubt*—but nor do I rule it out: I leave open that maybe I *shouldn't* be confident of *Thesis.*

Question: how should my attitudes toward *Thesis* and *Doubt* relate? *Thesis* is a claim about some subject-matter. *Doubt* is a claim about what opinion I ought to have about that subject-matter. Let's call my opinion about *Doubt* a **higher-order opinion**—an opinion about what opinion I should have. Since I am uncertain about what opinions I should have, I have *higher-order uncertainty.* Let's call my opinion toward *Thesis* a **first-order opinion**—an opinion about something other than what opinions I should have. Generalizing our question: how should my first-order and higher-order opinions relate? For example: if I become more confident that I shouldn't be confident of *Thesis*, should that lead me to be less confident of *Thesis*? Or: if I have a lot of higher-order uncertainty about how confident I should be in *Thesis*, can I nevertheless be fairly confident of it? I will not give a full answer to such questions—but I will take three steps toward one.

First step. Many have asked similar questions. But they have often framed it as a question of how one *body* of evidence—your *first-order evidence*—interacts with another body of evidence—your *higher-order evidence.*[2] My first claim:

> REFRAMING: We should reframe the question: Given your *total* evidence, how should your first- and higher-order *opinions* relate?

I defend REFRAMING by showing how to build a general framework for studying the relationship between first- and higher-order uncertainty (§1), and then putting it to work (§§2–3).

Second step. So reframed, our question is nontrivial:

> MODEST TRUISM: Your total evidence often warrants being uncertain what opinions your total evidence warrants, and (hence) being *modest*: uncertain whether you're rational.[3]

I defend the MODEST TRUISM by arguing that rational modesty—i.e. rational higher-order uncertainty—is needed to account for the epistemic force of disagreement (§2).

Third step. Many have pointed out that it seems irrational to believe that *my Thesis is true, but I shouldn't believe it.* The inferred explanation has been that your first-order opinions must "line up" with your higher-order opinions. Call this the **Enkratic**

---

[2]E.g. Feldman (2005); Christensen (2010a, 2016); Horowitz (2014); Schoenfield (2015, 2016); Sliwa and Horowitz (2015).

[3]I will assume a *single* normative notion that privileges certain opinions. I'll call them the opinions that "you *should* have," that "your (total) *evidence warrants*," or that "are *rational.*" If you think these normative notions come apart, please replace my expressions with your preferred, univocal one.

Intuition. Many theories defend (or presuppose) it as the answer to our question.[4] My final claim—my *Thesis*—is that the Enkratic Intuition is wrong:

> AKRATIC: If modesty is rational, so too is epistemic akrasia.

I defend AKRATIC by using the above framework to precisify the Enkratic Intuition and show that it is inconsistent with higher-order uncertainty (§3; cf. Titelbaum 2015).

That is the plan. Here is the picture. Higher-order uncertainty is pervasive and important. There is a general, tractable framework for studying it. Many open questions remain.

1 Two Problems

Recall that I'm confident of:

> *Thesis:* Epistemic akrasia can be rational.

But I also suspect that:

> *Doubt:* I should not be confident of *Thesis*.

In thinking about cases like this, the standard operating procedure is to make (something like) the following distinction:

(1) My *first-order evidence* about *Thesis* is the evidence that bears directly on it. (Example: my current arguments.)

(2) My *higher-order evidence* about *Thesis* is the evidence that bears *in*directly on it by bearing directly on claims like *Doubt*. (Example: the flaws in my past arguments.)

Authors making this distinction often presuppose that we can meaningfully speak of two distinct *bodies* of evidence—my first-order evidence, and my higher-order evidence.[5] Distinction made, the standard question goes something like this. Given my first-order evidence, I should have some opinion about *Thesis*. Now add my higher-order evidence. Should my opinion in *Thesis* change? If so, how? Since we are to imagine two interacting bodies of evidence, call this the **Two-Body Problem**.

My first claim is:

---

[4] E.g. Feldman (2005); Gibbons (2006); Christensen (2010b); Huemer (2011); Smithies (2012, 2015); Greco (2014); Horowitz (2014); Titelbaum (2015); Sliwa and Horowitz (2015); Littlejohn (2015); Worsnip (2015); Salow (2017).

[5] E.g. Feldman (2005); Christensen (2010a, 2016); Horowitz (2014); Schoenfield (2015, 2016); Sliwa and Horowitz (2015).

REFRAMING: We should reframe the question: Given your *total* evidence, how should your first- and higher-order *opinions* relate?

In other words: instead of two interacting *bodies* of evidence, we have two interacting *levels of opinions* warranted by a single, total body of evidence. We have a **Two-Level Problem**, not a Two-Body one.

I have no short, knock-down argument for REFRAMING. Instead, what I have to offer is (my own) confusion generated by the Two-Body Problem, and clarity generated by the Two-Level one. Perhaps you will share them.

Confusion first. One question: what exactly does it mean for a bit of evidence to bear "indirectly" on *Thesis*? In some sense, the claim that Jones has constructed an argument for *Thesis* does so. But this is not the sense has been meant in the higher-order evidence discussion, which focuses on the possibility of rational errors (Christensen 2010a)—sleep deprivation, hypoxia, irrationality pills, and the like. So perhaps evidence bears indirectly on *Thesis* when it bears on whether I've made a rational error in forming my opinion about *Thesis*? But suppose that I haven't yet formed any opinion about *Thesis*, and then the oracle informs me that *Doubt* is true. Surely this is still higher-order evidence, even though it says nothing about a rational error on my part.

Another question: how to these two bodies of evidence agglomerate? Suppose F is first-order evidence for q and H is higher-order evidence for q; what is the conjunction $F \wedge H$? It clearly bears directly on q, so it seems that it should be first-order evidence. But this means that when I go to base my beliefs about q on my first-order evidence, I will thereby base them on $F \wedge H$, bringing in the higher-order information H. So maybe those beliefs should be based on my *purely* first-order evidence. But what does it mean for a bit of evidence to be *purely*, *directly* about q? Consider q itself—surely this proposition is purely, directly about q if anything is. But consider the proposition:

Not-Known: My first-order evidence does not put me in a position to know $\neg q$.

Not-Known is a paradigm case of the sort of proposition that higher-order evidence about q works "through"—if a proposition p bears on *Not-Known*, it bears indirectly on q. But q *implies Not-Known*. Thus even q itself does not bear *purely* directly on q!

This is not meant to be clarifying. Nor is it meant to be a precise argument against the Two-Body Problem. What it is meant to be is an illustration of how easy it is to find oneself confused with this problem. My goal in the rest of the paper is to argue that the Two-*Level* Problem leads to a clearer framing of the questions and their potential answers.

1.1 Framing the debate

Assume that you have a single total body of evidence that determines what opinion you should have in any given proposition. Some of these propositions will be like *Doubt*—

claims about what opinions you should have, i.e. about what opinions your single, total body of evidence warrants having. Your opinions about such propositions are higher-order opinions. Other propositions will be like *Thesis*—claims that aren't about what opinions you should have. Your opinions about such propositions are first-order opinions. Here is an interesting question: how do the *first-order* opinions warranted by your total evidence relate to the *higher-order* opinions warranted by your total evidence?[6] For example, how confident of *Doubt* can my evidence warrant being before it necessarily warrants being less than confident of *Thesis*? Or: if we minimally change my evidence so that it warrants more confidence in *Doubt*, will it thereby warrant being less confident of *Thesis*? If so, how much?

We can state things more precisely. Let 'C' be a definite description for my actual degrees of belief—whatever they are. $[C(q) = t]$ is the proposition that I'm t-confident of q—it's true at some worlds, false at others. Let 'P' be a definite description for the credences I *should* have, given my (total) evidence. For simplicity, assume *unique precision*: my evidence always warrants a unique, precise probability function P.[7] $[P(q) = t]$ is the proposition that my (current, total) evidence warrants being t-confident of q. So at any given world w, there's a particular probability function that I ought to have—let 'P_w' be a rigid designator for the function initialized by w. (Unlike the definite descriptions 'P' and 'C', 'P_w' refers to a particular probability function whose values are fixed and known.) Since I can (rationally) be unsure which world I'm in, I can (rationally) be unsure which probability function my credences should match: if the open possibilities are $w_1, w_2, ...$ then I can leave open whether $[P = P_{w_1}]$ (the rational credence function is P_{w_1}) or $[P = P_{w_2}]$ (the rational credence function is P_{w_2}), or ....

With this notation in hand, here's how we can regiment my attitudes toward *Thesis* and *Doubt*. For simplicity, suppose I'm confident of q iff my credence in q is at least .7, and I leave open q iff my credence in q is nonzero. We can treat *Thesis* as a primitive proposition. On the other hand, *Doubt* is the proposition that I should not be confident of *Thesis*, i.e. that the rational credence in *Thesis* is less than .7. So *Doubt* = $[P(\textit{Thesis}) < .7]$. Thus my attitudes:

> I'm confident of *Thesis*: $[C(\textit{Thesis}) \geq .7]$,
> and I leave open that this confidence is rational: $[C(P(\textit{Thesis}) \geq .7) > 0]$.

> I leave open that I shouldn't be confident of *Thesis*: $[C(P(\textit{Thesis}) < .7) > 0]$.

---

[6] I'm certainly not the first to approach the issue in this way—see Williamson (2000, 2014, 2018); Christensen (2010b); Elga (2013); Lasonen-Aarnio (2015), and Salow (2017).

[7] It would be fairly straightforward to generalize the framework to drop this assumption. It's also worth noting that the models I use only presuppose int*ra*personal uniqueness: there is a uniquely rational credence function *for each agent*, given their information and standards of reasoning. For the (de)merits of these assumptions, see White (2005, 2009); Joyce (2010); Schoenfield (2014), and Schultheis (2017).

What's distinctive about my epistemic situation is that I'm unsure which opinions my (total) evidence warrants: I think maybe it warrants having a credence of at least .7, and maybe it warrants having a credence below .7. If we further assume that I'm sure of my *actual* credences—so $[C(C(\textit{Thesis}) \geq .7) = 1]$—it follows that I am unsure whether I'm rational; or, as I will say, I am *modest*.[8] For since I'm certain that I'm confident of *Thesis* and I leave open that I shouldn't be, I thereby leave open that I'm not rational: $\left[C\big(C(\textit{Thesis}) \neq P(\textit{Thesis})\big) > 0\right]$.

We might expect that if such higher-order doubts are warranted, then they should constrain my confidence in *Thesis*. The Two-Level Problem is whether, why, and to what extent this is so: how are rational opinions constrained by rational opinions about what opinions you should have? Notice that this is a question about how my higher-order doubts *should* affect my first-order opinions: it is a question about P (the credences I should have), *not* about C (my actual credences).

Stating the question more precisely: how is the value of $P(\textit{Thesis})$ modulated by the varying values of $P(P(\textit{Thesis}) = t)$ for various t? Using only the resources already specified, here are a host of natural answers that we could give to this Two-Level Problem:

> ACCESS INTERNALISM: $[P(q) = t] \rightarrow [P(P(q) = t) = 1]$
> If you should be t-confident of q, you should be certain that you should be t-confident of q.

> GRADED ACCESS: $[P(q) \geq t] \rightarrow [P(P(q) \geq t) \geq t]$
> If you should be at least t-confident of q, you should be at least t-confident that you should be at least t-confident of q.

> DEGREED JJ: $[P(P(q) \geq t) \geq s] \rightarrow [P(q) \geq ts]$
> If you should be at least s-confident that you should be at least t-confident of q, you should be at least $t{\cdot}s$-confident that q.

> REFLECTION: $P(q|P(q) = t) = t$
> Conditional on the rational credence in q being exactly t, you should adopt credence exactly t in q.

> SIMPLE TRUST: $P(q|P(q) \geq t) \geq t$
> Conditional on the rational credence in q being at least t, you should adopt a credence of at least t in q.

The goal of the Two-Level Problem is to assess principles like these for plausibility and tenability. Do they allow rational modesty? If so, do they nevertheless enforce plausible connections between first- and higher-order attitudes—or do they let such

---

[8]I follow Elga (2013) in the "modesty" terminology; note that it is orthogonal to the sense of "immodesty used in the epistemic utility theory literature (Lewis 1971)

attitudes split radically apart? The answers are often surprising. ACCESS INTERNALISM obviously rules out higher-order uncertainty, and implies each of the other principles (by trivializing them). Surprisingly, REFLECTION implies that you must always be certain of ACCESS INTERNALISM (as we will see in §3). GRADED ACCESS rules out most cases of higher-order uncertainty (see Williamson 2018). On the other hand, SIMPLE TRUST implies DEGREED JJ, and both of these principles allow massive amounts of higher-order uncertainty—meaning that SIMPLE TRUST is much weaker than REFLECTION.

It is not my aim here to explain or justify these particular assessments of these particular principles (cf. Dorst 2018). I mention them to give a sense of the terrain—for my aim is to explain why the Two-Level Problem is a fruitful and tractable strategy for exploring the notion of higher-order evidence.

To do that, I need to do three things. First, I need to address the foundational questions of how to model and interpret rational higher-order uncertainty (§1.2). Second, I need to argue that the solution to the Two-Level Problem is not the trivial one given by ACCESS INTERNALISM—that higher-order uncertainty is often rational (§2). Finally, I need to argue that the solution to the Two-Level Problem is not the obvious one given by the Enkratic Intuition or REFLECTION (§3).

1.2 Modeling it

We want to model a particular agent (say, me) at a particular time (say, now) who's uncertain about a particular subject matter (say, *Thesis*).

I should be uncertain about *Thesis*. How do we model that? By saying that I should match my opinions to a probability function that's uncertain of which world it's in—it assigns positive probability to *Thesis*-worlds and positive probability to ¬*Thesis*-worlds.

I should also be uncertain about whether *I should be confident of Thesis*. How do we model *that*? The same way. By saying that I should match my opinions to a probability function that's uncertain which world it's in—it assigns positive probability to *I-should-be-confident-in-Thesis*-worlds, and positive probability to *I-should-not-be-confident-in-Thesis*-worlds. That is, just as *Thesis* expresses a proposition, so too *I should be confident of Thesis* expresses a proposition. What we need is a systematic way to represent such propositions.

Here's how.[9] Let W be a (finite) set of epistemic possibilities that capture the distinctions relevant to my scenario. Propositions are modeled as subsets of W. Truth is modeled as membership, so q is true at w iff $w \in q$. Logical relations are modeled as set-theoretic ones, so: $\neg q = W - q$; $q \wedge r = q \cap r$; etc.

---

[9] I'm drawing on the probabilistic epistemic logic literature, though it usually assumes you know your own probabilities (cf. van Ditmarsch et al. 2015); some exceptions: Samet (1997); Williamson (2000, 2008, 2014); Lasonen-Aarnio (2015); Salow (2017).

'C' is a definite description for my *actual* degrees of confidence—whatever they are. It can be modeled as a function from worlds w to credence functions C_w—for simplicity, suppose C_w is always a probability function over W. (Note that while 'C' is a definite description that picks out different functions at different worlds, 'C_w' is a rigid designator for the credence function I have at world w.) Using it we can define propositions (subsets of W) about what I actually think. For any proposition $q \subseteq W$ and $t \in [0,1]$, let $[C(q) = t]$ be the proposition that I'm actually t-confident of q: $[C(q) = t] =_{df} \{w | C_w(q) = t\}$.[10]

'P' is a definite description for the degrees of confidence I *should* have—whatever they are. It too can be modeled as a function from worlds w to probability functions P_w over W, thought of as the credences I ought to have at w. What's crucial for modeling higher-order uncertainty is that we can use P it to define propositions about what I should think. For any proposition $q \subseteq W$ and $t \in [0,1]$, $[P(q) = t]$ is the proposition that I should be t-confident of q: $[P(q) = t] =_{df} \{w | P_w(q) = t\}$. Since we have identified facts about rational credences as propositions (sets of worlds), your (rational) credences are thereby defined for any higher-order claim about what credences you should have—that is, (rational) higher-order opinions fall right out of the model.

In sum, we can model my epistemic situation with a **credal-probability frame** $\langle W, C, P \rangle$ capturing the relevant possibilities (W), what I *actually* think in those various possibilities (C), and what I *should* think in those various possibilities (P).

I know that I should have credences that match P. I also—perhaps—know what my actual credences are. Higher-order uncertainty slips in because I may not know whether what I *actually* think (C) lines up with what I *should* think (P). If we assume that rational agents know their actual credences, such higher-order uncertainty can be rational iff there can be agents who are *in fact* rational—$[C = P]$—but who are *modest*: they are not certain that they are rational—$[C(C = P) < 1]$.

To get a grip on how this machinery works, let's construct a toy model of my case: I'm confident of *Thesis*, but I am uncertain whether I'm rational to be confident of *Thesis*. Suppose I know that I should either be .7 or .6 confident of *Thesis*. Letting T abbreviate *Thesis*, here is what we would like to say about my case:

(1) I should be sure that I should either be .7 or .6 confident of *Thesis*:
     $[P([P(T) = .7] \vee [P(T) = .6]) = 1]$.

(2) In fact I should be .7 confident of *Thesis*: $[P(T) = .7]$

(3) I should leave open that I should be .7, but also leave open that I should be .6:
     $[P(P(T) = .7) > 0]$ and $[P(P(T) = .6) > 0]$.

---

[10]Similar definitions apply to other claims about my confidence, e.g. that I'm more confident in q than r: $[C(q) > C(r)] =_{df} \{w | C_w(q) > C_w(r)\}$.

(4) I'm in fact .7 confident of *Thesis*, and I should be certain that I am: $[C(T) = .7]$ and $[P(C(T) = .7) = 1]$.

(5) My credences are in fact warranted by my evidence: $[C = P]$.

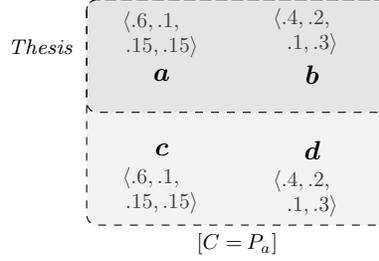Figure 1 is a credal-probability frame that makes (1)–(5) true at worlds a and c.

Figure 1: *Thesis* Uncertainty

There are four relevant epistemic possibilities: $W = \{a, b, c, d\}$. *Thesis* is true at worlds a and b, so *Thesis* $= \{a, b\}$ (hence $\neg$*Thesis* $= \{c, d\}$). Since I know my actual credences, C is a constant function: at each world w, C_w matches the credences that are rational at world a; $C_w = P_a$ for all w. (Indicated by the label for the shaded region covering all worlds.) The sequences next to each world w indicate the credences I should have in the various possibilities, in alphabetical order. So the '$\langle .6, .1, .15, .15 \rangle$' next to a indicates that $P_a(a) = .6$, $P_a(b) = .1$, $P_a(c) = .15$, and $P_a(d) = .15$. This in turn specifies the rational credences to have in any proposition by summing across worlds: $P_a(T) = P_a(\{a, b\}) = P_a(a) + P_a(b) = .7$. By our definitions, $[P(T) = .7] = \{w | P_w(T) = .7\} = \{a, c\}$—at a and c I *should* be .7 confident in my thesis—while $[P(T) = .6] = \{b, d\}$—at b and d I should be .6.

Here's the crucial point. At worlds a and c I should be .7 confident of *Thesis*. Yet at those worlds I should also assign positive credence to b and d—where I should instead be .6 confident of *Thesis*. This means I should have higher-order uncertainty: I should be .7 confident of *Thesis*, but I should leave open that I should instead be .6 confident of it. Precisely, (1)–(5) are true at worlds a and c for the following reasons:

(1) $[P(T) = .7] = \{a, c\}$ and $[P(T) = .6] = \{b, d\}$, so $([P(T) = .7] \vee [P(T) = .6]) = W$. So $[P([P(T) = .7] \vee [P(T) = .6]) = 1] = [P(W) = 1] = W$.

(2) $[P(T) = .7] = \{a, c\}$.

(3) Every world w is such that $P_w(\{a, c\}) > 0$ and $P_w(\{b, d\}) > 0$, so $[P(P(T) = .7) > 0]$ and $[P(P(T) = .6) > 0]$ are true everywhere.

(4) $[C(T) = .7]$ is true everywhere since $[C = P_a] = W$ and $P_a(T) = .7$. Since $[C(T) = .7] = W$, $[P(C(T) = .7) = 1] = [P(W) = 1] = W$.[11]

(5) $[C = P]$ is true at $\{a, c\}$ since $[C = P_a]$ is true everywhere and $[P = P_a] = \{a, c\}$.

This is the framework within which I'm proposing we study higher-order evidence: the framework of higher-order uncertainty, modeled using (credal-)probability frames. It provides the formal backbone to REFRAMING. Most of my argument will consist of putting it to work.

Before doing so, two final notes. First, it does not presuppose anything about how the warranted credences at various worlds are related to each other (for instance, they do not have to be recoverable by conditioning from a common prior). Thus it is consistent with the view that higher-order evidence should lead you to "bracket" some of your information (Christensen 2010a)—or in some other way provides counterexamples to conditionalization. Second, W is a set of *epistemic* possibilities. Thus there is no formal problem with using such a framework to model (higher-order) uncertainty about logic. If you are unsure whether an argument is valid, we can simply add epistemic possibilities where it is (isn't)—so long as we treat the claim that the argument is valid as an atomic proposition, no formal problems will arise. Difficult interpretive questions will arise, of course—but those exist for all approaches to modeling logically non-omniscient agents.

For an initial application, let me illustrate how we can use this sort of "total-evidence" framework to define notions that correspond fairly well to the intuitive ideas of what's warranted by your first- and higher-order evidence. Intuitively, the opinions warranted by my first-order evidence (my arguments) are simply the opinions that someone who was fully informed about evidential matters—who has no doubts about what was warranted by my evidence—would think. The reason I'm unsure what my first-order evidence warrants (and, therefore, what my total evidence warrants) is that I *do* have such higher-order doubts about evidential matters. Thus to determine what my first-order evidence warrants, we can ask: what opinions would I be warranted in having *if*

---

[11]Here is a very subtle point. At world b, the credence in *Thesis* warranted by my evidence is .6. Nevertheless, at b the credence warranted in the claim that my actual credence in *Thesis* is .7 is 1. That is, at b: $[P(T) = .6]$ and $[P(C(T) = .7) = 1]$. Though puzzling, this is correct. For in b I can tell what my actual credences are—I have overwhelming evidence that my credence in *Thesis* is in fact .7, thus my evidence warrants being certain of this claim. In b those credences are not warranted by my evidence—instead I should have .6 credence in *Thesis*. What this shows is that we can't understand P_b as giving the credences that *a rational agent* at b *would* have. Arguably, no rational agent could have credence .6 in *Thesis* while having my same (overwhelming) evidence that she has credence .7 in it—in adopting credence .6 in *Thesis* she would make it so that she had *different* evidence than I have (cf. Salow 2017). In short, P_b captures the opinions that are warranted (rational) *given my evidence*—not necessarily the opinions that *would* be warranted if I were to *conform* to my evidence, since in so conforming I may *change* my evidence (e.g. my evidence about my beliefs).

all my higher-order doubts were removed? That is, if I were to learn what the rational credence function was (i.e. what it was before I learned what it was), what would be the rational reaction to this information?[12]

Let's apply this thought to Figure 1. Notice that the credence function warranted at world a assigns .7 probability to *Thesis*—but it does so, in part, because it has higher-order uncertainty: it assigns .25 credence to being at b or at d, where a different credence in *Thesis* is rational. In other words, the rational .7 credence is modulated by higher-order doubts. At a, what would the rational opinions be if my higher-order doubts were removed? Let $\widehat{P}$ capture these opinions: $\widehat{\boldsymbol{P}}_{\boldsymbol{w}}(\cdot) =_{\boldsymbol{df}} \boldsymbol{P}_{\boldsymbol{w}}(\cdot | \boldsymbol{P} = \boldsymbol{P}_{\boldsymbol{w}})$, and $[\widehat{\boldsymbol{P}}(\boldsymbol{q}) = \boldsymbol{t}] =_{\boldsymbol{df}} \{\boldsymbol{w} | \widehat{\boldsymbol{P}}_{\boldsymbol{w}}(\boldsymbol{q}) = \boldsymbol{t}\}$ (cf. Stalnaker 2017). Since $\widehat{P}$ captures what the rational credences would be if higher-order doubts were removed, it can plausibly be understood as what my first-order evidence warrants. In Figure 1, $[P = P_a] = \{a, c\}$, so $\widehat{P}_a(T) = P_a(T | P = P_a) = \frac{P_a(T \wedge [P = P_a])}{P_a(P = P_a)} = \frac{P_a(a)}{P_a(\{a,c\})} = \frac{.6}{.75} = .8$. Hence $[\widehat{P}(\textit{Thesis}) = .8]$ is true at a and c, while a similar calculation shows that $[\widehat{P}(\textit{Thesis}) = .4]$ is true at b and d. So at a and c the first-order evidence strongly supports *Thesis* (my arguments are good), while at b and d it actually tells *against Thesis* (my arguments are bad). Yet both of these opinions are modulated by higher-order doubts to the more moderate opinions of .7 and .6.

We have the opinions warranted by your total evidence (P) and those warranted by your first-order evidence ($\widehat{P}$); what about the opinions warranted by your higher-order evidence? Intuitively, the opinions warranted by your total evidence should be "factorable" into the various possibilities you leave open for what your first-order evidence warrants, and your higher-order opinions about how likely those possibilities are to be actual. To see this, consider how we might alternatively represent Figure 1. There are two possibilities for what the first-order evidence warrants—$\{a, c\}$ and $\{b, d\}$. In this frame, each world agrees on the probability distribution *within* such cells: conditioned on $\{a, c\}$ or $\{b, d\}$, every P_w has the same distribution. The differences between the P_w are due to their distributions *across* such cells: the worlds in $\{a, c\}$ are split 75-25 between $\{a, c\}$ and $\{b, d\}$, while those in $\{b, d\}$ are split 50-50. Thus this frame can be equivalently represented using numbers *within* cells to indicate the first-order support there, and labeled arrows *between* cells to indicate the probability that the (total) evidence gives to being in each cell. That yields Figure 2.

In this picture, the credence in *Thesis* warranted by the total evidence at world a can be calculated by averaging the support of the two first-order-evidence cells, with weights determined by how confident P_a is of each cell: $P_a(T) = .75 \cdot .8 + .25 \cdot .4 = .7$. Similarly, $P_b(T) = .5 \cdot .8 + 5 \cdot .4 = .6$. In fact, the reason we can redraw Figure 1

---

[12]Careful here. If I learn the values of P, and P had higher-order uncertainty, then I learn something that P didn't know. Thus, as we're about to see, the rational reaction to learning the values of P may be different from P itself (Elga 2013; Hall 1994, 2004).
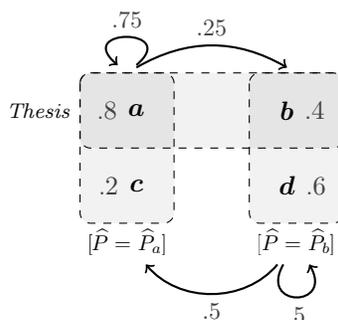
Figure 2: First- and Higher-Order Support

as Figure 2 is precisely because in Figure 1 this equality holds generally: the rational credence in q equals the rational expectation of the credence in q warranted by the first-order evidence. What do I mean? The rational expectation $\mathbb{E}_P[X]$ of a quantity X is a weighted average of the various possible values of X, with weights determined by how confident you should be in each. If you should be $\frac{1}{3}$ confident I have 3 hats and $\frac{2}{3}$ confident that I have 6 hats, then your rational expectation of my number of hats (the number you most expect to be *near* correct) is $\frac{1}{3}(3) + \frac{2}{3}(6) = 5$. Formally, $\mathbb{E}_P[X] = \sum_t (P(X = t) \cdot t)$. In the case at hand, the quantity we're estimating is the first-order support for q, $\widehat{P}(q)$, so $\mathbb{E}_P[\widehat{P}(q)] = \sum_t (P(\widehat{P}(q) = t) \cdot t)$. Given this, the principle that allows us to redraw Figure 1 as Figure 2 is:

HiFi: $P(q) = \mathbb{E}_P[\widehat{P}(q)]$

The rational credence in q ($P(q)$) equals the rational expectation of: the credence in q that's warranted by the first-order evidence ($\mathbb{E}_P[\widehat{P}(q)]$).

I call this principle HiFi because it captures the idea that the opinions warranted by your total evidence are factorable into your Higher-order expectations of your First-order evidence. In particular, when this principle holds it is natural to identify the opinions warranted by your *higher-order evidence* as simply the distribution across first-order-evidence cells that is warranted by your total evidence, i.e. $P(\widehat{P} = \widehat{P}_w)$ for various w. (This distribution is represented in the labeled arrows between cells in Figure 2.)

The upshot of this discussion is that for probability frames that validate HiFi, we can define well-behaved precisifications of the idea that some opinions are warranted by your "first-order evidence", and others are warranted by you "higher-order evidence."[13] Which frames do so?

HiFi turns out to be equivalent to the NEW REFLECTION principle proposed by Adam Elga (2013). NEW REFLECTION starts with the observation that if your evidence

---

[13]Admittedly, I have not told you what it means for a proposition to *be* first- or higher-order evidence. I have no idea how (or whether) that can be done.

warrants being uncertain of what your evidence warrants, then if you *learn*[14] what your evidence warrants, you have gained new information that was not already entailed by your evidence. (If $P(P = P_w) < 1$, then $P(\cdot | P = P_w)$ is more informed than $P(\cdot)$.) So what should you do when you learn what opinions your evidence warrants? Elga says: adopt the opinions that your evidence *would* warrant if it were to be updated with what you've just learned. Precisely:

> NEW REFLECTION: $P(\cdot | P = P_w) = P_w(\cdot | P = P_w)$
> Upon learning the opinions warranted by the evidence, react to this information in the way that you (now) know the evidence would warrant.

NEW REFLECTION sounds truistic. It is one way of making precise the idea that your opinions should be guided by your opinions about what your evidence warrants. And it is what allows us to "factor" your total evidence into first- and higher-order components:

Fact 1. *A probability frame* $\langle W, P \rangle$ *validates* HIFI *iff it validates* NEW REFLECTION.[15]

Nevertheless, there are objections to NEW REFLECTION (Lasonen-Aarnio 2015). It is not my goal here to defend the principle, but instead merely to show that it represents a choice-point in our ability to vindicate a version of the first/higher-order evidence distinction. (And to argue—below in §3.2—that NEW REFLECTION and HIFI should not be seen as the *solution* to the problem of higher-order evidence.)

This concludes my proposal for how to think about higher-order evidence—the details behind REFRAMING. The rest of the paper applies it. §2 defends the MODEST TRUISM that higher-order uncertainty is often rational, while §3 argues that the obvious principles for connecting first- and higher-order opinions do not succeed.

2 Rational Modesty

In this section I argue for:

> MODEST TRUISM: Your total evidence often warrants being uncertain what opinions your total evidence warrants, and (hence) being *modest*: uncertain whether you're rational.

Your evidence warrants higher-order uncertainty iff for some q and all t: $P(P(q) = t) < 1$. So long as you know what your actual opinions are, you have higher-order uncertainty iff you are modest, so I will treat modesty and higher-order uncertainty together.

---

[14]NEW REFLECTION is strictly about conditional beliefs, not learning. For ease of exposition I'll switch between talk of the two—but we could reformulate everything in terms of conditional beliefs.

[15]Proof in the Appendix; cf. Stalnaker (2017). A probability frame is a credal-probability frame without C. A frame *validates* a principle iff it makes the principle true at all worlds for all instantiations on which it is well-defined.

Isn't it obvious that we often do—and *should*—have such self-doubts? Intuitive cases abound. *Bias*: I'm inclined to think that Kim's job talk wasn't great; but, knowing the literature, I have good reason to suspect that I have implicit bias against her—I'm probably underappreciating her talk. *Impairment*: the answer to the test's "challenge problem" seems obvious; but I'm running on four hours of sleeping—I'm probably missing something. *Disagreement*: I thought the evidence supported the defendant's innocence; but *you* thought it supported his guilt—perhaps I've mis-assessed it. And so on.

Clean cases can also be found (Christensen 2010a; Elga 2013; Schoenfield 2016):

> HYPOXIA
>
> Flying your plane, you've done some reasoning and become confident that 10,000 feet is a safe altitude (*Safe*). Then over the radio you're told there's a good chance you're hypoxic, in which case your opinions may be slightly irrational. You know, given all this information, that you should be either *somewhat* or *fairly* confident of *Safe*. In fact you are fairly confident of *Safe*.

Isn't it obvious that in HYPOXIA you should be uncertain whether (1) your fair confidence is rational, or (2) you should instead be only somewhat confident in *Safe*? That is, isn't it obvious that you shouldn't be certain of what is warranted by your *total* evidence (including the radio announcement)?

From one perspective, it certainly seems so. Being rational is *hard*. Very often we don't live up to the challenge. We know this about ourselves. So very often we should think that maybe *right now* we're not living up to the challenge—we should be unsure what it's rational to think.

But from another perspective, to admit such rational higher-order uncertainty is to give up the game. For—the thought goes—*getting to the truth* is the hard part, and the job of epistemology is to provide us with in-principle-accessible rules that ensure we do the best we can. If we allow higher-order uncertainty, we will have to deny a form of this "in-principle-accessible" claim:

> ACCESS INTERNALISM: $[P(q) = t] \rightarrow [P(P(q) = t) = 1]$
> If you should be t-confident of q, you should be certain that you should be t-confident of q.

To deny this principle is to say that sometimes you are required to have an opinion even though you can't be sure that you are so required. This can seem unacceptable: failing to live up to requirements is grounds for criticism; how could you be legitimately criticized if you couldn't tell what was required of you? Those attracted to this line of thought will want a different way to think about our cases.

Let's focus on HYPOXIA. Suppose that in this context you're *fairly confident* iff your credence is .7, and you're *somewhat confident* iff your credence is .6. Then the natural

reading of the case is that you should be uncertain whether the rational credence in *Safe* is .6 or .7: $[P(P(Safe) = .7) > 0]$ and $[P(P(Safe) = .6) > 0]$. We can use the same model of me wondering about *Thesis* (from §1.2) to model you wondering about *Safe* (Figure 3). All of the above discussion applies equally well to this HYPOXIA case—you're

$Safe$	$\langle .6, .1,$	$\langle .4, .2,$
	$.15, .15 \rangle$	$.1, .3 \rangle$
	a	*b*
	c	*d*
	$\langle .6, .1,$	$\langle .4, .2,$
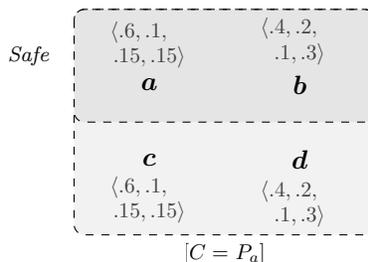|	$.15, .15 \rangle$	$.1, .3 \rangle$

$$[C = P_a]$$

Figure 3: Hypoxic Uncertainty

.7 confident of *Safe* and should be sure that you are, you are (and should) be unsure whether you should instead be .6, etc.

Question: is there a recipe for generating an internalist-friendly reading of cases like this? The main strategy I know of goes as follows.[16] It's intuitive to say, of a case like HYPOXIA, that "You should be uncertain of what you should think." If we interpret both those 'should's in the same way, then this says you should have higher-order uncertainty. But we needn't interpret them that way. Instead, we can interpret them as picking out different normative notions: there's (1) what you should think *given* your cognitive imperfections, and (2) what you 'should' think in the sense of what an *ideal* agent (with your evidence) would think. Thus the true reading of the sentence is: "You should (*given your imperfections*) be uncertain of what you should (*ideally*) think." Moreover, you should (ideally) know what you should (ideally) think; and you should (given your imperfections) know what you should (given your imperfections) think. Instead of higher-order uncertainty *within* a normative notion, these cases reveal first-order uncertainty *across* normative notions.

So far, so fair. But one more bit of explanation is needed: *Why* does each normative notion have no higher-order uncertainty? It's not too hard to get a sense for why you should ideally be certain of what you should ideally think—ideal agents are special, after all. But what explains why you should *given your imperfections* be certain of what you should *given your imperfections* think? The line of reasoning from above is just as intuitive. Properly accounting for our imperfections is *hard*. Very often we don't live up to the challenge. We know this about ourselves. So very often we should (given

---

[16]The strategy is inspired by Stalnaker (2017)—though he may not agree with my formulation of it.

our imperfections) think that maybe *right now* we're not properly accounting for our imperfections—we should (given our imperfections) be unsure what we should (given our imperfections) think. We should be modest.

An internalist may reply as follows. When we are uncertain of what we should (ideally) think, what we should (given our imperfections) do is to match our opinions to our *expectation* of what we should (ideally) think. Since we know this, we *do* know what we should (given our imperfections) think.

This strategy faces a dilemma. To illustrate, interpret "you should (ideally) have credence t" as "your *first-order evidence* warrants credence t", as defined in §1.2. On that definition, the first-order evidence warrants t-confidence in q ($\widehat{P}(q) = t$) iff the rational credence to have *once your higher-order doubts are removed* is t. What happens if we run the internalist reasoning using P and $\widehat{P}$? It goes as follows:

(1) You should (given your imperfections) be uncertain of what you should (ideally) think about *Safe*: $P(\widehat{P}(\textit{Safe}) = t) < 1$, for all t.

(2) You know that you should (given your imperfections) match your credence in *Safe* to your expectation of the credence $\widehat{P}(\textit{Safe})$ that you should ideally have.

(3) Therefore, you should (given your imperfections) be certain of what you should (given your imperfections) think: $[P(q) = t] \to [P(P(q) = t) = 1]$.

There is a mistake in this reasoning. Premise (2) can be interpreted in two ways. On one interpretation, it is false. On the other, it is true—but our frame validates it and (3) does not follow.

As we've seen, expectations of quantities are weighted averages of their possible values. Your *actual* expectations come from using your actual credences as the weights: $\mathbb{E}_C[X] = \sum_t (C(X = t) \cdot t)$. Premise (2) says that you should match your credence in *Safe* to your expectation of what the first-order evidence supports about *Safe*, i.e. to $\widehat{P}(\textit{Safe})$. But that premise can be interpreted in two ways, depending on whether the 'should' (represented with a '$\Box$') takes narrow- or wide-scope:

NARROW: $[\mathbb{E}_C[\widehat{P}(q)] = t] \to \Box[C(q) = t]$
If your expectation of the credence you should ideally have in q is *in fact* t, then you should (given your imperfections) have credence t in q.

WIDE: $\Box\big([\mathbb{E}_C[\widehat{P}(q)] = t] \to [C(q) = t]\big)$
You should (given your imperfections) be such that: if your expectation of the credence you should ideally have in q is t, then you have credence t in q.

NARROW says that whatever your actual expectation of the first-order evidence happens to be, that should determine your credence. If we assume that you know what your actual expectations are, this rules out modesty—as the internalist hoped. But NARROW

is false. Suppose we alter the description of HYPOXIA so that—for no reason at all—you expect the first-order evidence to warrant being certain of *Safe*. In this alternate case, would it follow that you should be certain of *Safe*? Of course not—what would follow is that you should change your expectation of the first-order evidence. So if we interpret Premise (2) as NARROW, (3) follows but (2) is false.

What's plausibly true is WIDE: you should be such that your actual credence lines up with your expectations of what the first-order evidence warrants. Given our assumption of unique precision, WIDE is equivalent to our familiar claim HIFI:

> HIFI: $P(q) = \mathbb{E}_P[\widehat{P}(q)]$
>
> The rational credence in q ($P(q)$) equals the rational expectation of: the credence in q that's warranted by the first-order evidence ($\mathbb{E}_P[\widehat{P}(q)]$).

As discussed above, Figure 3 validates HIFI. Thus this interpretation of premise (2) permits higher-order uncertainty: the expectations you should have ($\mathbb{E}_P[\cdot]$) depend precisely on what credences you should have (P)—therefore you cannot use knowledge of the former to gain knowledge of the latter.

Upshot: the "expectational" strategy for generating internalist-friendly readings of our cases does not succeed. Of course, this is not a refutation of ACCESS INTERNALISM. But it shifts the burden: the intuitive cases *do* put pressure on internalism.

Here, then, is how I see the dialectic. In many cases it's natural to think that you should be modest—and we have no internalist recipe for re-describing them. Nevertheless, internalists may make the principled claim that rational requirements must be accessible—that there must be *some* way to faithfully re-describe these cases. Nothing I've said so far is meant to dislodge this principled stand.

But the problem with principled stands is that they are brittle. Find a *single* exception, and the principle is shattered—the floodgates open. That is what I'll try to do. I'll argue that the epistemic force of peer disagreement cannot be accounted for without allowing rational higher-order uncertainty. With one case established, there's no reason to resist the natural hypothesis that rational higher-order uncertainty is *pervasive*.

2.1 Disagreement

I'll now argue that higher-order uncertainty is needed to make sense of the epistemic force of disagreement. First, the big picture.

Consider a situation in which you know that you and a peer have received independent, disjoint bodies of evidence about whether q. You are not confident of q, but then you discover that your peer *is* confident of q. All should agree that in this case you should increase your confidence in q. For you previously should've been unsure what your peer's (disjoint, independent) evidence supported. Your peer's opinion provides

you with evidence that her evidence supports q; thus you've received some evidence that *were you to pool your evidence*, the resulting more informed body of evidence would support q—you have received "evidence of evidence" for q (Feldman 2005). In general, you should defer to more informed bodies of evidence. So since you are now more confident that your pooled body of evidence supports q, you should increase your confidence in q.

Now consider a more standard peer disagreement case. You know that you and a peer have received the *same* body of evidence about whether q. You are not confident of q, but then you discover that your peer *is* confident of q. What should happen next? According to the higher-order uncertainty picture, the situation is precisely parallel. Since you previously should've been modest, you should've been unsure what your peer's (and your!) evidence supported. Your peer's opinion provides evidence that your (shared) evidence supported q; thus you've received some evidence that your evidence supported q—you have received "evidence of evidence" for q. In general, you should defer to what your evidence supports. So since you are now more confident that your evidence supported q, you should increase your confidence in q.

In short, we can use higher-order uncertainty to give a natural picture of the epistemic force of peer disagreement—one that is continuous with other widely acknowledged "evidence of evidence" effects. In fact, I'll argue that we *must*: if peer disagreement is to have the epistemic force that it's standardly taken to have, then higher-order uncertainty must be rational. I have two arguments.

The first begins with a highly circumscribed—yet paradigm—case of peer disagreement. The key features are these. First, you should be sure that you and your peer have the same relevant evidence with respect to some proposition q. Second, you should be sure that neither of you have any *non*-evidential connection to the truth.[17] My claim is that—in cases where these conditions hold—if learning that your peer disagrees with you should have *any* affect on your opinion about q, then you must have higher-order uncertainty.

To fix judgments, here's a concrete case:

> JUDGES
>
> You know all of the following. You and your peer Judy are judges who have served on the same court for years. You've both just heard the same case, in which you were presented with the same (relevant) total evidence. You have the same (relevant) standards of reasoning, have studied the same (relevant) cases, laws, precedents, and so on. Neither of you has any way to get to the truth, except by means of your evidence (no occult powers or

---

[17]This is to screen off the possibility that a your credences could be further evidence for q even after the we know what opinions your evidence warrants; (cf. Levinstein 2017; Titelbaum and Kopec 2017).

special intuitions). You each have gone to form your own opinions before convening.

Claim: in JUDGES, learning that Judy is less confident than you are that the defendant is liable (*Liable*) should lead you to change your credence that he's liable.[18] To get this verdict, you must have higher-order uncertainty.

First, the intuitive argument. Suppose you *don't* have higher-order uncertainty—you're certain of which opinion in *Liable* is warranted by your (total) evidence. Since you should be sure that you and Judy share relevant evidence, you should be sure that this opinion is likewise the opinion warranted by *Judy's* evidence. So since you're certain of what opinion Judy's evidence warrants—and you're certain she has no non-evidential connection to the truth—you will simply ignore her *actual* opinion. Contraposing: since you *shouldn't* ignore Judy's actual opinion, you *shouldn't* be certain of which opinion in *Liable* is warranted by your evidence.

Now the precise argument. What should your epistemic state look like, before you and Judy convene? Let 'P_y' be a definite description for the credences you (now) should have, and 'P_j' be one for the credences Judy (now) should have. Let 'C_j' be a definite description for Judy's actual credences. Let L be the proposition that the defendant is liable. Four premises.

First: since you know that you share evidence and standards of reasoning, you should be sure that the credence you should have in *Liable* is the same as the credence Judy should have:

Same: $[P_y\big(P_y(L) = P_j(L)\big) = 1]$

Second: since you and Judy are not perfect duplicates, you should leave open that her credence will be lower than yours:

Disagree: For all t: if $[P_y(L) = t]$, then $[P_y(C_j(L) < t) > 0]$

Third: since you know that Judy can only get to the truth via her evidence, if you learn[19] what credence she *should* have in *Liable*, then further learning what credence she *actually* has doesn't provide any evidence for or against *Liable*. So learning which credence Judy should have in *Liable* should screen off her actual credence:

Screening: For all t, s: $P_y\big(L\big|[P_j(L) = t] \wedge [C_j(L) = s]\big) = P_y(L|P_j(L) = t)$

---

[18]This is a very weak claim about the force of peer disagreement. It does not presuppose any view about how your credence should shift—it simply presupposes that it *should* shift. It is compatible with the Equal Weight View (Elga 2007), the Total Evidence View (Kelly 2010), "synergistic" views (Easwaran et al. 2016), and pretty much any formal proposal for pooling credences (cf. Brössel and Eder 2014; Pettigrew 2017). I believe only Right Reasons views (Titelbaum 2015) would deny it.

[19]Again, read my talk of "learning" as shorthand for conditional beliefs.

Finally: if you should have a given credence t, then upon learning that Judy's credence is *lower* than that, you shouldn't simply ignore her opinion.

Budge: For all t: if $P_y(L) = t$, then $P_y(L|C_j(L) < t) \neq P_y(L)$

These premises jointly imply that you should have higher-order uncertainty:

Fact 2. *If Same, Disagree, Screening, and Budge true at a world in a probability frame, so too are* $[P_y(P_y(L) = t) > 0]$ *and* $[P_y(P_y(L) \neq t) > 0]$ *for some t.*[20]

Upshot: to respect the epistemic force of peer disagreement in cases where you know that you and your peer share evidence and have no special access to the truth, higher-order uncertainty must be rational. If this is right, the principled internalist stand is shattered.

Internalists may respond by biting the bullet: grant that it is intuitive that in this case you shouldn't ignore Judy's disagreement, but insist that in fact you should. They may tell a debunking story: in *most* cases of peer disagreement, other dynamics are present which lead to rational conciliation—yet in the highly circumscribed cases described, these dynamics are gone.

I do not think such a response can work. The cases cannot be contained: to get the correct verdicts in *usual* cases of disagreement, higher-order uncertainty must be rational. This is my second argument.

Granted, *rarely* should you be certain that the credence you ought to have in q is *identical* to the credence your peer ought to have. However, *almost always* you should consider it possible that you and your peer will disagree in a way that implies that one of you was irrational. Maybe your opinions will very far apart (*too* far apart); maybe the manner in which you disagree will reveal that you were thinking very differently (*too* differently) about a piece of evidence; maybe something else.

Suppose you are on a jury with an equally smart, equally informed peer Pete. All the evidence has been presented, but you have not yet convened to share your opinions. Let g be the proposition that the defendant is *g*uilty. Let 'P^y' and 'P^p' be definite

---

[20]*Proof:* Given a probability frame $\langle W, P_y \rangle$, suppose Same, Disagree, Screening, and Budge are true at w. (I won't add functions P_j and C_j; assume that the relevant propositions obey the expected logical relations.) For reductio, suppose the consequent of Fact 2 is false; so for some t', $[P_y(P_y(L) = t') = 1]$ is true. Recalling that probability frames are finite, by finite additivity, Disagree implies that there are values $s_i < t$ such that $P_y(C_j(L) = s_i) > 0$. By total probability, $P_y(L|C_j(L) < t)$ is a weighted average of the values of $P_y(L|C_j(L) = s_i)$ (with some weights possibly 0); so to establish that $P_y(L|C_j(L) < t) = P_y(L)$ it will suffice to show that $P_y(L|C_j(L) = s_i) = P_y(L)$ for each s_i. Since $P_y(C_j(L) = s_i) > 0$, by our hypothesis that $P_y(P_y(L) = t') = 1$, we have $P_y(L|C_j(L) = s_i) = P_y(L|[P_y(L) = t'] \wedge [C_j(L) = s_i])$. By Same, this equals $P_y(L|[P_j(L) = t'] \wedge [C_j(L) = s_i])$. By Screening, this in turn equals $P_y(L|P_j(L) = t')$. By Same again, this equals $P_y(L|P_y(L) = t')$, which by hypothesis equals $P_y(L)$. It follows that $P_y(L|C_j(L) < t) = P_y(L)$, contradicting Budge.

descriptions for the opinions *you* and *pete* *should* have (before convening). Let 'C^y' and 'C^p' be definite descriptions for the opinions that *you* and *pete* *actually* have, before convening. Let *Disagree* be the proposition that one of your opinions (before convening) was irrational, i.e. $Disagree =_{df} ([C^p(g) \neq P^p(g)] \vee [C^y(g) \neq P^y(g)])$. Three premises.

First : before you convene with Pete, you should leave open that you two will disagree. (After all, you can't be certain that *Pete* is rational.)

Open: $P^y(Disagree) > 0$

Second: since we are responding to Access Internalism, we may safely assume that you should be certain of your actual credence that the defendant is guilty.

Actual: $[C^y(g) = t] \rightarrow [P^y(C^y(g) = t) = 1]$

Third: what should you think if you learn that you and Pete *do* disagree? You two are equally smart—hence, initially, equally likely to be (ir)rational. If you disagree, it follows that one of you *was* irrational. It would be arbitrary (and immodest) to assume that it must have been him. So upon learning *Disagree*, you should not be certain that Pete was irrational, nor certain that you were. Let 'P^y_D' be a definite description for the rational credences you should have upon learning that you and Pete disagree, i.e. $P^y_D(\cdot) =_{df} P^y(\cdot|Disagree)$. Then:

Uncertain: $P^y_D(C^p(g) \neq P^p(g)) < 1$ and $P^y_D(C^y(g) \neq P^y(g)) < 1$.

If Open, Actual, and Uncertain correctly describe your scenario, it follows that before you convene with Pete you should have higher-order uncertainty:

Fact 3. *If Open, Actual, and Uncertain are true at a world in a credal-probability frame, so too are $[P^y(P^y(g) = t) > 0]$ and $[P^y(P^y(g) \neq t) > 0]$ for some t.*[21]

If I am right that these premises correctly describe your scenario in *typical* cases of peer disagreement, it follows that such cases involve higher-order uncertainty.

---

[21] *Proof:* Given a credal-probability frame $\langle W, C^y, P^y \rangle$, suppose Open, Actual, and Uncertain are all true at some world. (Again, I won't formalize C^p and P^p; assume they obey the expected logical relations.) There will be some t for which $[C^y(g) = t]$ is true. We first show that $P^y(P^y(g) = t) > 0$. Notice that $P^y(Disagree \wedge [P^y(g) = C^y(g)]) = P^y(Disagree) \cdot P^y_D(P^y(g) = C^y(g))$. By Open, the first multiplicand is > 0. The second multiplicand equals $1 - P^y_D(P^y(g) \neq C^y(g))$. Since by Uncertain the subtracted term is < 1, it follows that $P^y_D(P^y(g) = C^y(g)) > 0$ as well. Combined, we have that $P^y(Disagree \wedge [P^y(g) = C^y(g)]) > 0$, and so $P^y(P^y(g) = C^y(g)) > 0$. Since by Actual and our supposition that $[C^y(g) = t]$ we have $P^y(C^y(g) = t) = 1$, it follows that $[P^y(P^y(g) = t) > 0]$ is true. Next we show that $[P^y(P^y(g) \neq t) > 0]$ is also true. By parallel reasoning (through Open and Uncertain), we have that $P^y(Disagree \wedge [P^p(g) = C^p(g)]) > 0$. Note that every world in which $Disagree \wedge [P^p(g) = C^p(g)]$ is true is one in which $[P^y(g) \neq C^y(g)]$; thus $P^y(P^y(g) \neq C^y(g)) > 0$. Since $P^y(C^y(g) = t) = 1$, we've established that $[P^y(P^y(g) \neq t) > 0]$, as desired.

I believe the only way to resist this argument is to deny Uncertain—claiming that in cases where you and a peer share similar evidence, if you learn that your peer disagrees with you (i.e. that one of you was irrational), then you must be certain that you were rational and they were not. This is a desperate move.

For one, we can make the case more extreme. Suppose that we re-run the scenario thousands of times and that in the cases where you *Disagree*, each one of you is (ir)rational equally often. Now we run the experiment again, and you discover that you disagree. Should you be certain that *this* time it was your peer who was irrational? That seems absurd. For two, unlike my first argument, the scenario described by this second argument is pervasive—it happens every time you should think you might fundamentally disagree with a peer about morality, religion, or philosophy.

If that's right, the internalist stand is shattered. Rational higher-order uncertainty is possible—in fact, pervasive.

3 Enkrasia

So far I have defended REFRAMING (that we should think of higher-order evidence in terms of higher-order uncertainty—as a Two-Level Problem) and a MODEST TRUISM (the Two-Level Problem is nontrivial—higher-order uncertainty can be rational). If this is right, the Two-Level Problem is well-formed and nontrivial.

But does it have a simple solution? Many have seemed to suggest so. They point out that the following states seem to be irrational: (1) believing that *my Thesis is true, but I shouldn't believe it*; (2) being confident that *my Thesis is true, but I shouldn't be confident of it*; and (3) believing *Thesis* while being agnostic on whether that belief is rational. The inferred explanation has standardly been that rationality requires your first-order opinions to "line up" with your higher-order ones—that your first-order opinions must be sanctioned (or, at least, *not disavowed*) by your higher-order opinions. Call this the **Enkratic Intuition**. Many theories of higher-order evidence have been built on top of it.[22] The Enkratic Intuition can be given a precise characterization within the higher-order-uncertainty framework. So if it is correct, our Two-Level Problem admits of a simple solution.

But it is not correct. Here I defend (cf. Titelbaum 2015):

> AKRATIC: If modesty is rational, so too is epistemic akrasia.

My strategy is as follows. I'll first argue that the Enkratic Intuition has a precise consequence for the relationship between your credence in q and your opinions about

---

[22]Including Feldman (2005); Gibbons (2006); Christensen (2010b); Huemer (2011); Smithies (2012, 2015); Greco (2014); Horowitz (2014); Sliwa and Horowitz (2015); Titelbaum (2015); Worsnip (2015); Littlejohn (2015); Rasmussen et al. (2016); Salow (2017), and perhaps Vavova (2014).

the rational credence in q. Then in §3.1 I'll show that this consequence is inconsistent leaving open the rationality of higher-order uncertainty. Since we often *should* leave open that higher-order uncertainty is rational, we should often be akratic. This does not show that (1)–(3) can be rational, of course. What it shows is that *if* they can't be, a different explanation is needed.

What does the Enkratic Intuition imply about the relationship between first- and higher-order credences? Suppose you're .5 confident that it'll rain tomorrow, yet you're symmetrically uncertain whether this credence is overconfident, just right, or underconfident:

$C(Rain) = .5$, while:
$$C(P(Rain) = .4) = .35$$
$$C(P(Rain) = .5) = .3$$
$$C(P(Rain) = .6) = .35$$

Then your .5 credence seems perfectly-well sanctioned by your higher-order opinion—the "pressure" from your higher-order beliefs to change your first-order credence is balanced.

Why is that? Here's a helpful metaphor (cf. Bertsekas and Tsitsiklis 2008, p. 83). The Enkratic Intuition suggests that to the degree you think the rational credence is higher than yours, that should "pull" your credence upward; and to the degree you think the rational credence is lower than yours, that should "pull" your credence downward. So imagine a bar labeled '0' at one end and '1' at the other is resting on a fulcrum:

Now imagine attaching a block on each spot $t \in [0, 1]$ along the bar with weight proportional to $C(P(q) = t)$—your credence that the rational credence in q is t. This will tip the scale:
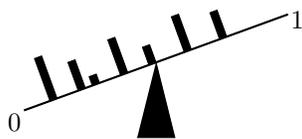
Question: where would you have to put the fulcrum to *balance* the scale, so that it's not leaning left or right? (Where would you have to put your credence to balance the pull of your higher-order doubts?) Answer: you must place it at the *center of gravity c*:

23
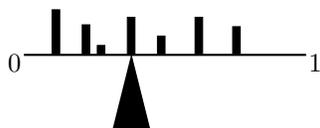
This is the point at which the sum of the torques from the weights to the left of the fulcrum ($t < c$) is equal to the sum of the torques from the weights to the right ($c < t$). The torque of a block weighing $C(P(q) = t)$ at a distance $|t - c|$ is simply the weight times the distance: $C(P(q) = t) \cdot |t - c|$. A bit of algebra shows that this means that $c = \sum_t \left(C(P(q) = t) \cdot t \right)$. This formula should be familiar: it is the mathematical expectation of $P(q)$, calculated relative to C. So our metaphor leads to the plausible conclusion that your credence in q is enkratic only if it equals your expectation of the rational credence in q. Recalling (from §2) that $\mathbb{E}_C[X] =_{df} \sum_t \left(C(X = t) \cdot t \right)$ is your actual expectation of X, we have:

ENKRATIC: Your credence in q is enkratic only if $C(q) = \mathbb{E}_C[P(q)]$

I claim that ENKRATIC captures what the Enkratic Intuition requires of your degrees of belief.[23] Of course, ENKRATIC is a precisification of a intuitive principle that has primarily been motivated by appeal to the irrationality of certain *outright* beliefs, so there is a risk of a terminological impasse here. My point is simply that the intuitive *motivation* for the claim that believing *p but I shouldn't believe it* is irrational is equally motivation for ENKRATIC. Thus if ENKRATIC must be rejected—as I'll argue it must— then we cannot look to the Enkratic Intuition to solve the Two-Level Problem.

Suppose ENKRATIC is correct. Then enkrasia is a rational norm only if the *rational* credence in q ($P(q)$) equals the *rational* expectation of the rational credence in q ($\mathbb{E}_P[P(q)]$):

RATIONAL ENKRASIA: $P(q) = \mathbb{E}_P[P(q)]$

Notice that this is *not* HIFI, for we are here calculating the expectation of the rational credence—not the expectation of the credence warranted by the first-order evidence. (More on this in §3.2.) RATIONAL ENKRASIA is a consequence of the simple REFLECTION principle discussed above: that upon learning that the rational credence in q is t, you should adopt credence t in q; $P(q|P(q) = t) = t$. So what do RATIONAL ENKRASIA and REFLECTION require?

3.1 Enkratic? Immodest.

They require that you be certain that you should be immodest.

If RATIONAL ENKRASIA is true, that's because it is a structural requirement that helps constrain the rational response to higher-order doubts. So if enkrasia is a rational requirement and modesty is rational, it should be possible to *know* that enkrasia is a rational requirement while being modest. (We shouldn't endorse a view on which

---

[23]This conclusion is not original to me (Christensen 2010b; Sliwa and Horowitz 2015; Rasmussen et al. 2016; Salow 2017).

the *only* times you can be modest are when you are uncertain whether you should be enkratic.)

Problem: it turns out that if an agent knows that they should be enkratic, they must be certain that they should be immodest. Recall that a principle is *valid* on a probability frame iff it is true at all worlds for all instantiations of its free variables. If you know RATIONAL ENKRASIA, it must be valid on the probability frame that represents your epistemic situation. What do such frames look like? Letting Sq be the proposition that you should be *S*ure of q, i.e. $Sq =_{df} [P(q) = 1]$:

Fact 4 (Samet 1997). *If a probability frame validates* RATIONAL ENKRASIA*, then it validates* $S\big([P = \pi] \leftrightarrow S[P = \pi]\big)$.[24]

$[P = \pi] \leftrightarrow S[P = \pi]$ is true for all instantiations of π at a world w iff whatever the rational credence function is at w, you should be certain *that* it's the rational credence function—iff you should be immodest. So to say the frame validates $S\big([P = \pi] \leftrightarrow S[P = \pi]\big)$ is to say that you should be certain that you should be immodest. So if RATIONAL ENKRASIA is correct, anyone who knows that it is must be certain that they should be immodest. But they *shouldn't* be certain that they should be immodest. So RATIONAL ENKRASIA is incorrect.

Why does RATIONAL ENKRASIA rule out higher-order uncertainty? The basic reason was discussed in conjunction with Elga's (2013) NEW REFLECTION principle above. When higher-order doubts are rational, then learning what the rational credences are provides new evidence, and so *changes* the rational credences (Elga 2013). Metaphorically, RATIONAL ENKRASIA enjoins you to aim at a moving target.

To see why, consider the following case. You and Selena share evidence but disagree: Selena is self-confident—she's .9 confident that she's rational—while you are modest— you're 50-50 on whether she or you is rational. Schematically, if s is the set of possibilities where Selena is rational, y is the set where you are, and 'C^y' is a definite description for your actual credences, we have Figure 4.

$$s \qquad\qquad y$$
$$\langle .9, .1 \rangle \qquad\qquad \langle .5, .5 \rangle$$
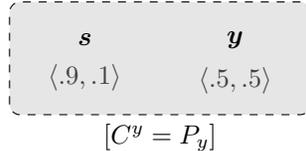$$[C^y = P_y]$$

Figure 4: Simple modesty

This frame illustrates why higher-order uncertainty is incompatible RATIONAL EN-KRASIA. If you're rational, then the rational credence that Selena is rational ($P(s)$) is

---

[24]This result is due to Samet (1997) (although he has a different intended interpretation of P). In the Appendix I give a less mathematically-involved proof than his.

.5. This is equivalent to the rational expectation of the truth-value $T(s)$ of s (0 or 1): at y, $P_y(s) = \mathbb{E}_{P_y}[T(s)] = 1 \cdot P_y(s) + 0 \cdot P_y(y) = .5$. In contrast, the rational expectation of the *rational credence* in s is: $\mathbb{E}_{P_y}[P(s)] = .9 \cdot P_y(P(s) = .9) + .5 \cdot P_y(P(s) = .5) = .9 \cdot P_y(s) + .5 \cdot P_y(y) = .7$. This expectation is higher than your expectation of the truth-value of s due to the fact that the rational credence is affected by higher-order doubts. You think it's .5 likely that Selena's rational—in which case the rational credence of .9 is slightly below the truth-value of 1; but you also think it's .5 likely that you're rational—in which case the rational credence of .5 is well above the truth-value of 0. These two divergences are asymmetric, so they do not cancel out—which is why your credence of .5 that Selena is rational is below your expectation of .7 for the rational credence. The crucial point that given higher-order uncertainty, you should *not* be trying to get your credence as close to the rational credence as you can—for sometimes the rational credence is modulated by higher-order doubts.

This is subtle, but can become intuitive. What is not intuitive—but what Fact 4 shows—is that the example generalizes completely: there is *no* way to "balance" your higher-order doubts to respect the requirements of RATIONAL ENKRASIA, short of the trivial case of higher-order certainty. I do not know of a way of making this result seem intuitively obvious. And I take that fact to be evidence that higher-order uncertainty is subtle, and that we do well to explicitly test our principles in a model theory like that of probability frames.

3.2   A Reply?

Perceptive readers sympathetic to the Enkratic Intuition may wonder: if the problem with "trying" to get close to the rational credence is that it's plagued with higher-order doubts, why don't we reformulate the enkratic requirement to aim at what your *first-order* evidence supports? Doing so would yield HiFi:

> HiFi: $P(q) = \mathbb{E}_P[\widehat{P}(q)]$
>
> The rational credence in q ($P(q)$) equals the rational expectation of: the credence in q that's warranted by the first-order evidence ($\mathbb{E}_P[\widehat{P}(q)]$).

As discussed above, *this* principle permits higher-order uncertainty. Why isn't *it* the proper precisification of the Enkratic Intuition?

Because it does not explain the cases that motivate that intuition. Granted, it *does* explain why attitudes like being confident that *it'll rain, but my first-order evidence doesn't support that* are irrational. But it *doesn't* explain why the following attitudes are irrational: (1) being confident that *it'll rain, but my <u>total</u> evidence doesn't support that* or (2) being very confident that *it'll rain, but my total evidence warrants being very confident it won't*. If any attitude is epistemically akratic, these ones are. Yet HiFi allows them. Example: two sycophants, Sybil and Phan, are each confident that the

other person is the rational one. Sybil is .9 confident that Phan is rational, while Phan is .9 confident that Sybil is. If these opinions could be rational, we'd have a probability frame like Figure 5 (s is the possibility where Sybil is rational, and p is the possibility where Phan is). Since both P_s and P_p agree on everything when they update on the

$$
\begin{array}{cc}
\boldsymbol{s} & \boldsymbol{p} \\
\langle .1, .9 \rangle & \langle .9, .1 \rangle
\end{array}
$$

Figure 5: Sycophants

claim that one of them is rational, this frame validates NEW REFLECTION and HIFI. Yet it gives rise to a paradigm case of akrasia:

$P_s(p \wedge [P(\neg p) \geq .9]) \geq .9$

At s they should be very confident that *Phan is rational but we should be very confident that he's not.*

Such an attitude looks akratic, if anything does. Any principle that allows it—like HIFI—cannot capture the Enkratic Intuition.

The upshot is clear: the Enkratic Intuition cannot form the foundation of a theory of higher-order uncertainty. We need some *other* principle to explain the irrationality of such radical splits between first- and higher-order attitudes.

4 Proposal

I've defended three claims. First is REFRAMING: the problem of higher-order evidence is best formulated as one of higher-order *uncertainty*—as a Two-Level Problem about how first- and higher-order opinions should relate. Second is MODEST TRUISM: such higher-order uncertainty is often rational—the solution to the Two-Level Problem is not trivial. Third is AKRATIC: the Enkratic Intuition is incorrect—the solution to the Two-Level Problem is not straightforward.

Where does this leave us? There are many reasons to want a general, strong theory of higher-order uncertainty. Such a theory would provide a foundation to the epistemologies of disagreement, debunking, and self-doubt. It would do so by formulating law-like principles connecting rational first- and higher-order opinions. It would use a framework like ours to both show that such principles are tenable and illustrate their consequences.

I have proposed such a theory elsewhere (Dorst 2018). But the framework proposed here is compatible with many, many alternatives. I hope to have done enough to show that it provides fruitful terrain—terrain that is well worth exploring.[25]

Appendix

Fact 1. *A probability frame* $\langle W, P \rangle$ *validates* HIFI *iff it validates* NEW REFLECTION.

Proof. We prove a LEMMA: in any frame, $[P = P_x] = [\widehat{P} = \widehat{P}_x]$. If $y \in [P = P_x]$, then $P_y = P_x$, so $\widehat{P}_y = P_y(\cdot | P = P_y) = P_x(\cdot | P = P_x) = \widehat{P}_x$, so $y \in [\widehat{P} = \widehat{P}_x]$. If $y \notin [P = P_x]$, then since $\widehat{P}_y(P = P_y) = 1$, $\widehat{P}_y(P = P_x) = 0$ while $\widehat{P}_x(P = P_x) = 1$, so $\widehat{P}_y \neq \widehat{P}_x$, so $y \notin [\widehat{P} = \widehat{P}_x]$.

Now suppose that NEW REFLECTION is valid. Taking an arbitrary world w and proposition q, since $\{[P = P_x]\}$ forms a partition:

$$
\begin{aligned}
P_w(q) &= \sum_{P_x} P_w(P = P_x) \cdot P_w(q | P = P_x) & \text{(total probability)} \\
&= \sum_{P_x} P_w(P = P_x) \cdot P_x(q | P = P_x) & \text{(NEW REFLECTION)} \\
&= \sum_{P_x} P_w(P = P_x) \cdot \widehat{P}_x(q) & \text{(definition)} \\
&= \sum_{P_x} P_w(\widehat{P} = \widehat{P}_x) \cdot \widehat{P}_x(q) \;=\; \mathbb{E}_{P_w}[\widehat{P}(q)] & \text{(LEMMA, DEFINITION)}
\end{aligned}
$$

So HIFI is valid.

For the converse, suppose that NEW REFLECTION is *not* valid, so there is a w, q, x such that $P_w(q | P = P_x) \neq P_x(q | P = P_x)$. Consider the proposition $q \wedge [P = P_x]$. $P_w(q \wedge [P = P_x]) = P_w(P = P_x) \cdot P_w(q \wedge [P = P_x] | P = P_x) =$

$$
P_w(P = P_x) \cdot P_w(q | P = P_x) \tag{1}
$$

Meanwhile, $\mathbb{E}_{P_w}[\widehat{P}(q \wedge [P = P_x])] = P_w(P = P_x) \cdot \widehat{P}_x(q \wedge [P = P_x]) =$

$$
P_w(P = P_x) \cdot P_x(q | P = P_x) \tag{2}
$$

Since by hypothesis $P_w(q | P = P_x) \neq P_x(q | P = P_x)$, it follows that $(1) \neq (2)$, and hence that $P_w(q \wedge [P = P_x]) \neq \mathbb{E}_{P_w}[\widehat{P}(q \wedge [P = P_x])]$, so HIFI is not valid. $\qquad \square$

Fact 4 (Samet 1997)**.** *If a probability frame validates* RATIONAL ENKRASIA, *then it validates* $S([P = \pi] \leftrightarrow S[P = \pi])$.

I will write $\mathbb{E}_w[P(q)] =_{df} \sum_{x \in W} (P_w(x) \cdot P_x(q))$ for the values of the expectation of $P(q)$ at w. Given a probability frame $\langle W, P \rangle$, there is an induced binary relation $\boldsymbol{R}$ such that wRx iff $P_w(x) > 0$ and $R_w = \{x | P_w(x) > 0\}$.

Lemma 4.1. *If* $\langle W, P \rangle$ *validates* RATIONAL ENKRASIA, *then R is transitive: if wRx and xRy then wRy.*

Proof. Suppose wRx and xRy but $w\cancel{R}y$. Since xRy, $P_x(y) > 0$. Since wRx, $P_w(x) > 0$. Therefore $\mathbb{E}_w[P(y)] \geq P_w(x) \cdot P_x(y) > 0$. Since by hypothesis $w\cancel{R}y$, $P_w(y) = 0$, contradicting RATIONAL ENKRASIA at w. □

Lemma 4.2. *If $\langle W, P \rangle$ validates* RATIONAL ENKRASIA*, then R is shift-reflexive: if wRy, then yRy.*

Proof. Suppose the frame validates RATIONAL ENKRASIA, so by Lemma 4.1 R is transitive. Suppose for reductio that wRy but $y\cancel{R}y$. Since R is transitive, for all $z_i \in R_y$, $P_{z_i}(R_y) = 1$. And since $y \notin R_y$ but $P_y(R_y) = 1$ and $P_x(y) > 0$, we have

$$\mathbb{E}_w[P(R_y)] \geq \sum_{z_i \in R_y} (P_w(z_i) \cdot P_{z_i}(R_y)) \; + \; P_w(y) \cdot P_y(R_y)$$
$$= \sum_{z_i \in R_y} P_w(z_i) \cdot 1 \; + \; P_w(y) \cdot 1 \; > \; \sum_{z_i \in R_y} P_w(z_i) = P_w(R_y)$$

Contradicting RATIONAL ENKRASIA at w. □

Lemma 4.3. *If $\langle W, P \rangle$ validates* RATIONAL ENKRASIA*, then R is shift-symmetric: if wRy and yRz, then zRy.*

Proof. Suppose the frame validates RATIONAL ENKRASIA, and so by Lemmas 4.1 and 4.2 R is transitive and shift-reflexive. Suppose for reductio that wRy and yRz but $z\cancel{R}y$. By transitivity, all $z_i \in R_z$ are such that $P_{z_i}(R_z) = 1$. By shift-reflexivity, $P_y(y) > 0$ and $z \in R_z$, so $P_y(R_z) > 0$. Finally, $y \notin R_z$. Combining these facts:

$$\mathbb{E}_y[P(R_z)] \geq \sum_{z_i \in R_z} (P_y(z_i) \cdot P_{z_i}(R_z)) \; + \; P_y(y) \cdot P_y(R_z)$$
$$= \sum_{z_i \in R_z} P_y(z_i) \cdot 1 \; + \; P_y(y) \cdot P_y(R_z) > \sum_{z_i \in R_z} P_y(z_i) = P_y(R_z)$$

Contradicting RATIONAL ENKRASIA at y. □

Lemma 4.4. *If $\langle W, P \rangle$ validates* RATIONAL ENKRASIA*, then for all $w \in W$: if wRy then $P_y(P = P_y) = 1$.*

Proof. Suppose the frame validates RATIONAL ENKRASIA, so by Lemmas 4.1 and 4.3, R is transitive and shift-symmetric. Suppose for reductio that wRy but $P_y(P = P_y) < 1$. By transitivity, if yRz and zRx, then yRx; equivalently: if yRz then $R_z \subseteq R_y$. By shift-symmetry, if yRz, then zRy; so by transitivity $R_y \subseteq R_z$ as well. Combined: if yRz, then $R_y = R_z$.

Since $P_y(P = P_y) < 1$, there must be a proposition q such that $P_y(P(q) = t) < 1$ for all t. Since W is finite, there is a set $T = \{t_1, ..., t_n\}$ such that for all t_i, $P_y(P(q) = t_i) > 0$, with at least two distinct $t_i \neq t_j$ in T. Relabel so that $t_1 < t_2 < ... < t_n$. There must

be some $z \in R_y$ such that $P_z(q) = t_n$. By the above reasoning, $R_z = R_y$, meaning T is also the set of values s such that $P_z(P(q) = s) > 0$. Then:

$$\mathbb{E}_z[P(q)] = \sum_{t_i \in T} \left(P_z(P(q) = t_i) \cdot t_i \right.$$

$$= \sum_{t_i < t_n} P_z(P(q) = t_i) \cdot t_i \ + \ P_z(P(q) = t_n) \cdot t_n$$

This is a weighted average with the highest possible value t_n. Since there are at least two values in T, the left summand has positive weight; thus the average is $\mathbb{E}_z[P(q)] < t_n$. Since $P_z(q) = t_n$, this contradicts RATIONAL ENKRASIA at z. $\qquad\square$

Fact 4 is an immediate consequence of Lemma 4.4.

References

Bertsekas, Dmitri P and Tsitsiklis, John N, 2008. *Introduction to Probability*. Athena Scientific, second edition.

Brössel, Peter and Eder, Anna-Maria A, 2014. 'How to resolve doxastic disagreement'. *Synthese*, 191(11).

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

———, 2010b. 'Rational Reflection'. *Philosophical Perspectives*, 24:121–140.

———, 2016. 'Disagreement, Drugs, etc.: From Accuracy to Akrasia'. *Episteme*.

Dorst, Kevin, 2018. 'Evidence: A Guide for the Uncertain'.

Easwaran, Kenny, Fenton-Glynn, Luke, Hitchcock, Christopher, and Velasco, Joel D, 2016. 'Updating on the Credences of Others: Disagreement, Agreement, and Synergy'. *Philosophers' Imprint*, 16(11).

Elga, Adam, 2007. 'Reflection and Disagreement'. *Noûs*, 41(3):478–502.

———, 2013. 'The puzzle of the unmarked clock and the new rational reflection principle'. *Philosophical Studies*, 164(1):127–139.

Feldman, Richard, 2005. 'Respecting the Evidence'. *Philosophical Perspectives*, 19(1):95–119.

Gibbons, John, 2006. 'Access Externalism'. *Mind*, 115(457):19–39.

Greco, Daniel, 2014. 'Could KK be OK?' *Journal of Philosophy*, 111(4):169–197.

Hall, Ned, 1994. 'Correcting the Guide to Objective Chance'. *Mind*, 103(412):505–517.

———, 2004. 'Two Mistakes about Credence and Chance'. *Australasian Journal of Philosophy*, 82(1):93–111.

Horowitz, Sophie, 2014. 'Epistemic Akrasia'. *Noûs*, 48(4):718–744.

Huemer, Michael, 2011. 'The Puzzle of Metacoherence'. *Philosophy and Phenomenological Research*, 82(1):1–21.

Joyce, James M., 2010. 'A Defense of Imprecise Credences in Inference and Decision Making'. *Philosophical Perspectives*, 24(1):281–323.

Kelly, Thomas, 2010. 'Peer disagreement and higher order evidence'. In Alvin I Goldman and Dennis Whitcomb, eds., *Social Epistemology: Essential Readings*, 183–217. Oxford University Press.

Lasonen-Aarnio, Maria, 2015. 'New Rational Reflection and Internalism about Rationality'. In Tamar Szabó Gendler and John Hawthorne, eds., *Oxford Studies in Epistemology*, volume 5, 145–171. Oxford University Press.

Levinstein, Ben, 2017. 'Permissive Rationality and Sensitivity'. *Philosophy and Phenomenological Research*, XCIV(2):343–370.

Lewis, David, 1971. 'IMMODEST INDUCTIVE METHODS*'. *Philosophy of Science*, 38(1):54–63.

Littlejohn, Clayton, 2015. 'Stop Making Sense? On a Puzzle about Rationality'. *Philosophy and Phenomenological Research*, To Appear:1–19.

Pettigrew, Richard, 2017. 'Choosing for Changing Selves'.

Rasmussen, Mattias Skipper, Steglich-Petersen, Asbjørn, and Bjerring, Jens Christian, 2016. 'A Higher-Order Approach to Disagreement'. *Episteme*, to appear.

Salow, Bernhard, 2017. 'The Externalist's Guide to Fishing for Compliments'. *Mind*, To appear.

Samet, Dov, 1997. 'On the Triviality of High-Order Probabilistic Beliefs'. https://ideas.repec.org/p/wpa/wuwpga/9705001.html.

Schoenfield, Miriam, 2014. 'Permission to Believe: Why Permissivism is True and What it Tells Us About Irrelevant Influences On Belief'. *Nous*, 48(2):193–218.

———, 2015. 'A Dilemma for Calibrationism'. *Philosophy and Phenomenological Research*, 91(2):425–455.

———, 2016. 'An Accuracy Based Approach to Higher Order Evidence'. *Philosophy and Phenomenological Research*, To Appear.

Schultheis, Ginger, 2017. 'Living on the Edge: Against Epistemic Permissivism'. *Mind*, to appear.

Sliwa, Paulina and Horowitz, Sophi, 2015. 'Respecting *all* the evidence'. *Philosophical Studies*, 172(11):2835–2858.

Smithies, Declan, 2012. 'Moore's paradox and the accessibility of justification'. *Philosophy and Phenomenological Research*, 85(2):273–300.

———, 2015. 'Ideal Rationality and Logical Omniscience'. *Synthese*, 192(9):2769–2793.

Stalnaker, Robert, 2017. 'Rational Reflection, and the Notorious Unmarked Clock'.

Titelbaum, Michael, 2015. 'Rationality's Fixed Point (or: In Defense of Right Reason)'. In Tamar Szabó Gendler and John Hawthorne, eds., *Oxford Studies in Epistemology*, volume 5, 253–292. Oxford University Press.

Titelbaum, Michael G and Kopec, Matthew, 2017. 'Plausible Permissivism'. Oxford University Press.

van Ditmarsch, Hans, Halpern, Joseph Y, van der Hoek, Wiebe, and Kooi, Barteld, 2015. *Handbook of Epistemic Logic*. College Publications.

Vavova, Katia, 2014. 'Confidence, Evidence, and Disagreement'. *Erkenntnis*, 79:173–183.

White, Roger, 2005. 'Epistemic Permissiveness'. *Philosophical Perspectives*, 19(1):445–459.

———, 2009. 'Evidential Symmetry and mushy credence'. *Oxford Studies in Epistemology*, 161–186.

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

———, 2008. 'Why Epistemology Cannot be Operationalized'. In Quentin Smith, ed., *Epistemology: New Essays*, 277–300. Oxford University Press.

———, 2014. 'Very Improbable Knowing'. *Erkenntnis*, 79(5):971–999.

———, 2018. 'Evidence of Evidence in Epistemic Logic'. In Mattias Skipper and Asbjørn Steglich-Petersen, eds., *Higher-Order Evidence: New Essays*, volume To appear.

Worsnip, Alex, 2015. 'The Conflict of Evidence and Coherence'. *Philosophy and Phenomenological Research*, To Appear:1–42.