Evidence: A Guide for the Uncertain

Kevin Dorst

November 2016

Abstract

Truism: You should treat the evidence as a guide—learning what it supports should systematically affect your beliefs. But we are not omniscient about our own evidence. *Truism:* Higher-order uncertainty can be rational—we can get non-trivial higher-order evidence about what our evidence supports. *The Problem:* Our truisms are, it seems, inconsistent—higher-order uncertainty threatens to debunk the guiding role of evidence. *The Project:* Explain how evidence *is* (always) guiding, even though higher-order uncertainty *is* (usually) rational. I first argue that standard "Reflection" principles are too strong—they trivialize higher-order uncertainty—while Elga's (2013) *New* Reflection principle is too weak—it allows puzzles of higher-order evidence to proliferate. But the failures of these principles motivate a new principle, **Trust**, which formalizes the platitude that "What the evidence supports is likely to be true." I show that Trust (1) allows higher-order uncertainty, (2) banishes our puzzles, (3) is characterized by an elegant class of models which (4) have a natural interpretation and (5) provide a systematic way to model cases while avoiding puzzles. But what if Trust is too strong—or too weak? We can offer a *proof*, of sorts, that it's not. The puzzles of higher-order evidence can be unified as failures of the **value of evidence** in the sense made famous by I.J. Good (1967)—failures of the platitude that evidence makes your beliefs more accurate and your decisions more wise. I then show that Trust is an epistemic characterization of the value of evidence.

Keywords: Higher-order evidence; epistemic akrasia; value of evidence; probabilistic epistemic logic; reflection principles.

1 A Problem, a Project

On my way to the airport, I get a call: "It's likely the plane isn't safe." "According to what?" "Uncle Ron." Click. I don't worry about Uncle Ron—and neither should you. Ring, ring—another call. "It's likely the plane isn't safe." "According to what?" "The evidence." Screech! I *do* worry about the evidence—and so should you.

Truism: You should treat the evidence as a guide. Learning what it supports should systematically affect your beliefs. Our question: How so? A simple answer is the Certainty View: you should always be certain of what your evidence supports. Were it true, this answer would support a "Reflection" principle that would guarantee that you should treat the evidence as a guide.

But the Certainty View is false: we are not omniscient about what we should think; we can be informed or misled about it. *Case:* disagreement. Is the plane safe? The (peer) engineers share evidence but disagree. Lois thinks they should have *Lo*w credence that it's safe; Hiedi thinks they should have *Hi*gh credence. Mil is unsure—she thinks maybe Lois is right; maybe Hiedi is; maybe neither—so she has *Mi*ddling credence. Two points. (1) Mil has *higher-order uncertainty*: she's uncertain about what their evidence supports—uncertain whether she should have low, middling, or high credence. But (2) Mil may be perfectly rational. With complicated evidence and disagreeing peers, sometimes you *should* be uncertain of what you should think—engineering is hard, after all. *Truism:* Higher-order uncertainty can be rational; we can get nontrivial higher-order evidence about what our evidence supports.

The Problem: Our truisms are, it seems, inconsistent. Higher-order uncertainty threatens to debunk the guiding role of evidence. From 10,000 feet, here's why. If higher-order evidence is possible, *misleading* higher-order evidence is possible. So take a case where you have (misleading) evidence that your evidence doesn't support p. You should believe as your evidence supports (let's say), so you should believe *my evidence doesn't support p*. But since that evidence is misleading, what it supports is false: in fact your evidence *does* support p—so you should believe p too. Thus, it seems, you should be epistemically akratic: believing the conjunction *p, but my evidence doesn't support it*. That means you *shouldn't* treat your evidence as a guide after all, for you should think that conforming to it will lead you to miss out on a truth—namely, p. Upshot: the guiding role evidence is in tension with the rationality of higher-order uncertainty.

How to respond? Here the literature divides. *Bridgers*[1] think that although higher-order uncertainty can be rational, there is some rational bridge between first- and higher-order attitudes. At a first pass, Bridging seems obviously correct. But I and others have argued elsewhere that the existing proposals don't succeed—they are either too strong or too weak ([XXX], Lasonen-Aarnio 2015). Two reactions to these failures. *Splitters*[2] argue that since higher-order uncertainty is possible, there is *no rational connection* between first- and higher-order attitudes. Akrasia can be rational; you can sometimes

[1] Feldman (2005); Gibbons (2006); Elga (2007, 2013); White (2009); Christensen (2010a,b, 2016); Huemer (2011); Vavova (2014, 2016); Horowitz (2014); Schoenfield (2015, 2016); Sliwa and Horowitz (2015); Littlejohn (2016); Worsnip (2016); Das (ms).

[2] Williamson (2000, 2014); Lasonen-Aarnio (2010, 2014, 2015); Coates (2012); Hazlett (2012); Wedgwood (2012); Weatherson (ms).

expect the evidence to mislead you. In contrast, *Mergers*[3] argue that since akrasia is *ir*rational, there is *no rational separation* between first- and higher-order attitudes. This is the Certainty View: higher-order uncertainty is irrational.

The motivations for Splitting and Merging come from the failures of Bridging. Everyone should agree that *if* Bridging succeeds—if there's a principled center of gravity that respects *both* our truisms—then that's the way to go. I'm here to tell you that there is.

The Project: Explain how evidence *is* (always) guiding, even though higher-order uncertainty *is* (usually) rational. The theory is simple—for that we need a slogan. *Trust the evidence.* The explanation is longer—for that we need a paper. The central idea is that evidence plays its guiding role because *what the evidence supports is likely to be true.* I will show that this idea is motivated by previous proposals, permits higher-order uncertainty, banishes puzzles of higher-order evidence, and in fact characterizes the **value of evidence** in the sense made famous by I.J. Good (1967).

Plan: §2 shows that "Reflection" principles trivialize higher-order uncertainty. §3 explains how we can avoid such surprises by using models of probabilistic epistemic logic. Models in hand, §4 endorses Elga's (2013) diagnosis of *why* Reflection is too strong—but goes on to show that his proposed *New* Reflection is too weak, for it allows puzzles of higher-order evidence to proliferate. §5 presents my theory—**Trust**—as the goldilocks principle. §6 unifies our puzzles as failures of the value of evidence, and uncovers a big coincidence.

2 A Tried Theory

Suppose Mil discovers that it's rational to have .7 credence the plane is safe, given her evidence. Upon learning this, how confident should she be that the plane is safe? Natural answer: .7. Generalizing: conditional on the evidence supporting p to exactly degree t, you should be exactly t-confident of p. Formalizing (van Fraasen, 1984; Christensen, 2010b):

$$\textbf{Reflection:}\quad P^i(p|P^k(p)=t)=t \qquad (k\geq i)$$

Here P^i and P^k are evidential probabilities from bodies of evidence i and k. k must be *at least* as informed as i, so they may be identical ($k=i$) or k may be more informed ($k>i$). Reflection, then, says that upon learning that a body of evidence *at least as informed as your own* supports p to exactly degree t, you should be exactly t-confident of p. It has the ring of a truism—how could it be wrong?

Reflection is provably inconsistent with higher-order uncertainty—that's how. This is a consequence of a little-known theorem from Dov Samet (1997). Letting S^ip be the proposition that you should be *S*ure of p given evidence i (S^ip iff $P^i(p)=1$), we have:

[3]Smithies (2012, 2015, ms); Greco (2014); Titelbaum (2015); Salow (2016, ms).

Theorem* 2.1 (Samet)**.** *A (finite) general probabilistic frame validates Reflection only if it validates* $S^i\big([P^i(p) = t] \leftrightarrow S^i[P^i(p) = t]\big)$.[4]

The biconditional $[P^i(p) = t] \leftrightarrow S^i[P^i(p) = t]$ says that however confident you should be in p, you should be sure that you should be exactly that confident—higher-order uncertainty is irrational. Adding an S^i on the front says that you must be certain of this fact. Upshot: Reflection implies that either you must always be certain of an obvious falsehood, or higher-order uncertainty is always irrational. This is a triviality result. Reflection cannot capture the guiding role of evidence.

We want to figure out why this happens—and in §4 we will. But first we need to learn from our mistakes. Reflection has been widely used and discussed for decades—often in the context of higher-order probability.[5] Rarely has it been noticed that it trivializes the subject matter.[6] But there's a principled way to avoid such triviality surprises: define a model theory to check the satisfaction conditions of our proposed principles.

3 A Guide's Guide

Probability is hard—our intuitions are often surprised. Higher-order probability is nuts—our intuitions are all but useless. If we're to build a theory of it, we need something to guide them. We need models of *probabilistic epistemic logic*. As I explain in Appendix A, this formalism yields a tractable way to model the intricacies of cases like our running example:

> **The Engineers.** Lois, Mil, and Hiedi are engineers tasked with determining whether the plane is safe. They share a bunch of complicated evidence, and know that at least one of them will respond as they should—someone always does. They form opinions by using their evidence to settle questions that affect how likely the plane is to be safe, e.g. "How old is the engine?" Suppose they only disagree about the answer to one such question—namely, "Do the controls handle smoothly?" Hiedi is sure they do, so she has high (.9) credence the plane's safe. Lois is sure they don't, so she has low (.5) credence it is. Mil is on the fence: she thinks maybe Hiedi's got their evidence right; maybe Lois has; maybe neither—maybe their evidence doesn't settle whether the controls handle smoothly. On the one hand Mil is inclined to think the

[4] "General probabilistic frames" are the most general models needed to study higher-order probability, defined at the end of Appendix A. "Validates" means the formula holds at all worlds for all p, t, i. This result combines Samet's theorems 3 and 5, noting that his conditional can be strengthened to the biconditional above.

[5] Skyrms (1980, 1990); Gaifman (1988); Christensen (2010b); Sliwa and Horowitz (2015); Roush (2016); Rasmussen et al. (ming).

[6] Exceptions: Williamson (2000, 2014); Elga (2013); Lasonen-Aarnio (2015).

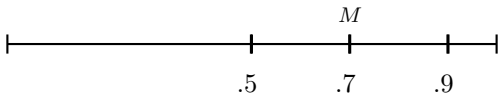
> controls *aren't* smooth. But on the other, Hiedi's more of an expert than Lois is. Thus Mil averages out to middling (.7) credence the plane is safe. Moreover, she has higher-order uncertainty, for she thinks maybe Hiedi (or Lois) is right that she should have high (or low) credence—or maybe they're wrong, in which case she *should* have middling credence. Alas. But she's not too worried, for tomorrow they are going to talk to Eve the expert, who will tell them who responded rationally.

Details in Appendix A; here's what you need to know. A **probabilistic frame** models a single agent—say, Mil—in a particular epistemic scenario, and consists of four components: $\langle W, R^1, R^2, \mu \rangle$. W captures the relevant possibilities and propositions—say, whether the plane is safe or not. R^1 and R^2 capture two different bodies of evidence—say, Mil's evidence today and tomorrow, respectively. μ captures background degrees of evidential support—say, Mil's (rational) standards of reasoning. *Crucial fact:* propositions about the evidence can be identified within W, just like propositions about the plane. $\boldsymbol{S^i p}$ is the set of worlds where Mil should be Sure of p, given the evidence in R^i. $[P^i(p) = t]$ is the set of worlds where evidence R^i makes it exactly t-likely that p. Thus we can model uncertainty about evidence: just as Mil can be uncertain today what the evidence tomorrow will support, so too can she be uncertain today what her evidence *today* supports—she can have higher-order uncertainty. For instance, perhaps she should be .4 confident that she should be .9 confident the plane is safe—or perhaps not: $[P^1(P^1(\mathit{safe}) = .9) = .4]$ is true at some worlds, false at others. Upshot: we can use probabilistic frames to model cases, formulate puzzles, and test principles. Let's get to it.

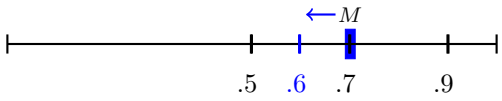
4 A True Theory...

Our framework in place, recall Reflection: upon learning that evidence at least as informed as your own supports p to exactly degree t, you should be exactly t-confident of p: $P^i(p|P^k(p) = t) = t$. Why is this too strong? Consider the (higher-order) case where $i = k$. As Adam Elga (2013) points out, once we allow higher-order uncertainty, the evidence may be uncertain that it supports p to exactly degree t. In that case, *learning* that $[P^i(p) = t]$ gives your evidence *new* information. And—quite generally—getting new information can change what's rational to think. Slogan: *Learning what you should think can* change *what you should think.*

Example: the engineers. Mil is rational but has higher-order uncertainty. She has credence .7 the plane is safe, but is unsure whether she should have .5 or .9 instead. So initially her evidence leaves open the following possible rational credences:



How confident should she be upon learning that her .7 credence was supported by her evidence, i.e. $[P^1(safe) = .7]$? Reflection would say '.7,' but let's think. This gives her new information—namely, that *she* is rational; Lois and Hiedi are not. And recall that she was initially inclined to think the controls *weren't* smooth, but her respect for Hiedi led her to raise her credence to .7. Now that she knows that Hiedi got this one wrong—that she should trust her own judgment—she should *lower* her credence that the plane's safe:



As is easy to check, at $\{a, b, c, d\}$ in Figure 7 (Appendix A) Reflection fails: $[P^1(safe|P^1(safe) = .7) = .6]$.

Upshot: in treating the evidence as a guide, we need to allow that it may have higher-order uncertainty. Elga (2013) proposes that if we learn something about the evidence, we need to make sure to *give* it this new information before we defer to it. Slogan: *When you learn about the evidence, respond as you know the evidence would.*

Elga offers the following formalization. Conditional on some particular function π being the rational credence function, your credence in p should equal π's credence in p *conditional on* π *being rational* (written $[P^k = \pi]$).[7] Formalizing:

$$\textbf{New Reflection:}\quad P^i(p|P^k = \pi) = \pi(p|P^k = \pi) \qquad (k \geq i)$$

As is easy to check, this principle yields the verdict we wanted for Mil in Figure 7: upon learning that she's rational, Mil should have .6 credence the plane is safe.

4.1 ...but Misguiding

New Reflection is true. But it can't be *all* that's true, for it doesn't guarantee that evidence is a guide. We show that New Reflection permits three increasingly bizarre puzzles of higher-order evidence. How? Probabilistic epistemic logic! We draw a probabilistic frame that validates New Reflection but permits our puzzles.

There are a variety of cases that can be used to illustrate this; here I'll use a particularly simple one. To be clear: I am not endorsing the following description of the case—I'm using it as a reductio.

[7] As with all propositions about the rational credences, we can identify this proposition as a set of worlds in our probabilistic frame. π *is rational* is simply the set of worlds where the rational credence function equals π: $\boldsymbol{[P^k = \pi] =_{df} \{w|P^k_w = \pi\}}$ (cf. Lasonen-Aarnio, 2015).

> **The Unmarked Clock.** Tim owns an unmarked clock with an hour-hand that can occupy one of twelve positions. Being a trickster, he sets it to a random position every day—but you're onto him. Later today you'll walk past his office and catch a glimpse of the clock, attempting to figure out where it's pointing.

The example is unproblematic. Here's the paradoxical description (cf. Williamson, 2014). Suppose the evidence you'll receive from your glimpse depends solely on where the hand is pointing. If it's at a given position, you should be sure it's within some "margin for error" around that position, with size determined by your reliability—say, ± 1. Before your glimpse you are completely unsure of which of the 12 positions it will occupy. If (say) it's pointing at 2, after your glimpse you should have $\frac{1}{3}$ credence in each of positions 1, 2, and 3. We can model this with the probabilistic frame in Figure 1. There are 12 possibilities in W. R^1 is trivial ($R^1 = W \times W$), so it's not drawn.

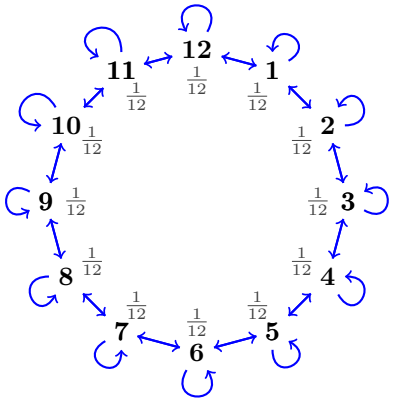


Figure 1: The Unmarked Clock

The blue arrows represent R^2: what will be consistent with your evidence in various possibilities, after your glimpse. The faded fractions represent μ: your background $\frac{1}{12}$ credence in each possibility.

New Reflection permits this description of the case:[8]

Proposition 4.1. *New Reflection is validated by The Unmarked Clock.*

In fact, Elga (2013) *designed* New Reflection to permit it. This was a mistake. The Unmarked Clock, so described, is a paradox.

First puzzle: Improbable Knowing. Interpreting S^i as knowledge, this puzzle was the focus of Williamson's (2014) original clock case. At each world there is a proposition p such that you should be sure of *p, but it's unlikely I should be sure of it.* Formalizing:

[8] See Appendix B for all proofs.

Improbable Knowing: $S^2(p \wedge [P^2(S^2p) < \frac{1}{2}])$

Example: if $p = \{1, 2, 3\}$, then $S^2p = \{2\}$ and $[P^2(Sp) \leq \frac{1}{3}] = \{1, 2, ..., 12\}$, so at 2 $S^2(p \wedge [P^2(S^2p) \leq \frac{1}{3}])$ is true.

Given Improbable Knowing, you shouldn't treat your evidence as a guide. For since you should be sure of p but confident the evidence will make you *less* sure of it, you should expect that conforming to the evidence will pull you away from a truth—namely, p. Strike one.

Second puzzle: Misguided Evidence. Recall our case of akratic beliefs: believing p *but I shouldn't believe it.* Suppose for the sake of argument that we adopt a Lockean theory of belief, wherein you should believe something iff it's sufficiently likely on your evidence—above some threshold T. Then akratic beliefs amount to being at least T-confident of p *I shouldn't be T-confident of p.* That's exactly what happens here. At each world there's a proposition p such that you should be confident of p *but I shouldn't be so confident of it.* Formalizing:

Misguided Evidence: $P^2\big(p \wedge [P^2(p) < t]\big) \geq t$

Example: if $p = odd = \{1, 3, 5, ..., 11\}$, then $p \wedge [P^2(p) \leq \frac{1}{3}]$ is true at 1 and 3, so $\big[P^2(p \wedge [P^2(p) \leq \frac{1}{3}]) \geq \frac{2}{3}\big]$ is true at 2.

Given Misguided Evidence, you shouldn't treat your evidence as a guide. For since you should be confident of p *but I shouldn't be confident of it*, the open possibilities where you should become less confident in p are precisely those where it's true. Again, you should expect that conforming to the evidence will pull you away from a truth—namely, p. Strike two.

Third puzzle: Self-Effacing Evidence.[9] There is a single proposition p such that at every world you should be *certain*—both before and after your glimpse—that the evidence is going to mislead you with respect to p. This certainty is true, safe, sensitive, etc.—it's *known.* You *know* that conforming to the (total) evidence will point you in the wrong direction with respect to p. Formalizing:

Self-Effacing Evidence: $S^i(p \leftrightarrow [P^2(p) < \frac{1}{2}])$
$S^i(\neg p \leftrightarrow [P^2(p) > \frac{1}{2}])$

Example: let $p = odd = \{1, 3, ..., 11\}$. Every odd possibility leaves open two even possibilities, and every even possibility leaves open two odd ones. Thus the biconditionals *it's odd iff I should be confident it's even* $(p \leftrightarrow [P^2(p) < \frac{1}{2}])$ and *it's even iff I should be confident it's odd* $(\neg p \leftrightarrow [P^2(p) > \frac{1}{2}])$ are true at every possibility. Since you're certain (you know) that you're in some possibility or other, these biconditionals are certain (known) to arbitrary iterations, both before and after you're glimpse.

[9] A version of this puzzle is anticipated by (Horowitz, 2014), though she bites the bullet and accepts it. I argue elsewhere that she shouldn't [XXX].

Given Self-Effacing Evidence, you *definitely* shouldn't treat your evidence as a guide. You know it's an *anti*-guide! You know that credence .5 in *odd* is more accurate than your evidence is. If you're offered a bet in favor of *odd* and a bet in favor of *even*, then you know that deciding as you *ought* to decide (using the credences warranted by the evidence) will lead you to take the wrong bet and lose money. Upshot: whatever it takes to treat the evidence as a guide, in a case of Self-Effacing Evidence you shouldn't. Strike three—New Reflection is out. It does not fully capture the guiding role of evidence.

Why? New Reflection requires that when you learn what the evidence supports, you use it to respond to this information. But unlike Reflection, it does not constrain what the evidence thinks about the evidence—it allows the evidence to distrust itself. This is what happens in the clock. Whenever $\pi(\mathit{odd})$ is high, $\pi(\mathit{odd}\big|P^2 = \pi)$ is low—the evidence knows that it supports *odd* only if *odd* is false.

5 A Trustworthy Theory

Puzzles proliferate, principles fail. What are we to do? *Trust the evidence.* Take it as a guide to the truth. Of course, the evidence can be misleading. But it's not *usually* misleading. It's *likely* that what the evidence supports is true. How likely? That depends on the strength of the evidence: strong evidence is almost never misleading; weak evidence is (almost) often so. Generalizing: supposing the evidence makes it $\left|\begin{smallmatrix}\text{fairly}\\ \text{quite}\\ \text{very}\end{smallmatrix}\right|$ likely that p, it's $\left|\begin{smallmatrix}\text{fairly}\\ \text{quite}\\ \text{very}\end{smallmatrix}\right|$ likely that p. Formalizing:

$$\textbf{Naive Trust:}\quad P^i(p\big|P^k(p) \geq t) \geq t \qquad (k \geq i)$$

Conditional on evidence at least as informed as your own making it t-likely that p, it's t-likely that p. Picturesquely: *most* of the open possibilities that make p probable are ones that make p true.

That's a first pass at the theory. But Naive Trust is naive. It requires that you trust the evidence's *actual* verdicts, but not it's *conditional* ones. It allows that adding new information could lead us to *dis*trust our (new) evidence. But we don't trust the evidence by happy chance—it *just is* the optimal handler of information. Generalizing: conditional on any q, if the evidence (given q) supports p, it's likely that p. Formalizing:

$$\textbf{Trust:}\quad P^i_q(p\big|P^k_q(p) \geq t) \geq t \qquad (k \geq i)$$

Conditional on evidence at least as informed as your own making it t-likely that p given q, it's t-likely that p given q. That's the principle—the theory. The rest is explanation.[10]

[10]Recall that $[P^i_q(p) = t]$ is a claim about the conditional probability $[P^i(p|q) = t]$. Fully written out, Trust is $P^i(p|q \wedge [P^k(p|q) \geq t]) \geq t$. Note: Naive Trust is the special case with $q = p \vee \neg p$.

Everyone's first reaction is that Trust must be asymmetric. It's not. Plugging in $p = \neg r$ and $t = 1 - s$ yields $P^i_q(r|P^k_q(r) \leq s) \leq s$. If the evidence makes it likely that p, it's likely that p; and if the evidence makes it *un*likely that p, it's *un*likely that p. No asymmetry here.

Everyone's second reaction is that Trust must (therefore) entail Reflection. It doesn't. (Why not? See below.)

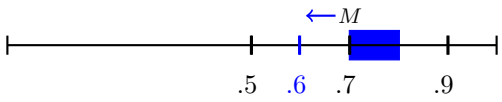
Everyone's third reaction is that—even so—any *motivation* for Trust must equally well be a motivation for Reflection, right?

Wrong. In fact, the idea behind *New* Reflection generalizes to motivate Trust. Recall the New Reflection slogan: *When you learn about the evidence, respond as you know the evidence would.* Now contrast two sorts of things you might learn about the evidence:

(1) The evidential support for p falls *within some range:* $[l \leq P^k(p) \leq h]$.
(2) The evidential support for p falls *above some threshold:* $[P^k(p) \geq t]$.

(2) is the sort of condition that appears in Trust, while (1) is what appears in Reflection—with the special case being $l = h$. There is a crucial difference between the two: you don't know how the evidence will respond when it learns (1), but you *do* know how it will respond when it learns (2)—you know it'll go up.

Here's why. When you learn (1) that $P^k(p)$ falls within some range—with upper and lower bounds—you don't know whether learning this will raise or lower the evidence's estimate of the rational credence in p.[11] That means you don't know whether the evidence will raise or lower its estimate *for* p. Since for all you know it could have started on the border of the $[l, h]$ range, upon learning what you've learned the evidential support might well fall *outside* the $[l, h]$ range. Example: Mil has .7 credence the plane is safe. But upon learning that the rational credence is between .7 and .8, she can infer that *she* (not Hiedi or Lois) is rational. So—as we've already seen—her credence drops to .6, outside the $[.7, .8]$ range:

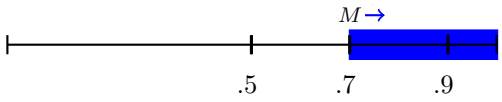


In contrast, when you learn (2) that $P^k(p)$ falls above some threshold t, you know that this will *raise* the evidence's estimate of the rational credence in p.[12] So you know the evidential probability *started* at least t and that its estimate of the rational credence went *up*. Question: could this cause it to lower it's probability for p? If it did, then upon raising its estimate of the rational credence in p, the evidence would lower it's support

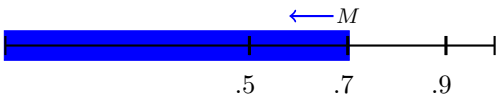
[11] Where $E^i[P^k(p)]$ is the mathematical expectation of $P^k(p)$ given P^i: it could be that $E^i\left[P^k(p)\middle|l \leq P^k(p) \leq h\right]$ is higher or lower than $E^i[P^k(p)]$.
[12] Theorem: $E^i\left[P^k(p)\middle|P^k(p) \geq t\right] \geq E^i[P^k(p)]$.

for p. That'd mean the evidence thinks the evidential probability is not a guide to the truth! But it *can't* think that—that's our first truism. Upshot: you can infer that upon learning what you've learned (namely, $P^k(p) \geq t$), the evidence's probability for p will go up. Since you know it started at least t and went up, you know that upon learning what you've learned it will wind up at least t. And now New Reflection's slogan kicks in: respond as you know the evidence would! Since you know the evidence will be at least t in p upon learning $[P^k(p) \geq t]$, *you* should be at least t upon learning this. Trust holds.

Example: If Mil learns $[P^1(p) \geq .7]$, she's only ruled out Lois's low credence—so she'll now be split between her original judgment and Hiedi's:



And similarly with the "downward-facing" version of Trust. If she learns $[P^1(p) \leq .7]$, she's only ruled out Hiedi's high credence—so she'll now be split between her original judgment and Lois's:



Moreover, we can use Mil's case to see *why* Trust doesn't entail Reflection. We know $P^1(p|P^1(p) \geq .7) \geq .7$ and $P^1(p|P^1(p) \leq .7) \leq .7$. Why doesn't it follow that combining both conditions requires Mil to have credence *exactly* .7, as with Reflection? Take it in stages. Suppose Mil learns the rational credence is at most .7. Following Trust, she moves to credence $P^+(p) = P^1(p|P^1(p) \leq .7) \approx .57$. Now she learns further that the original rational credence was also *at least* .7, i.e. that $[P^1(p) \geq .7]$. Should she now move to exactly .7?

No. Pay attention to the superscripts. The above argument showed that if she has credences P^1 and learns *only* that $[P^1(p) \geq .7]$, she should have at least .7. But she *doesn't* have credence P^1 any more—she has P^+, since she also knows that $[P^1(p) \leq .7]$. And she should respond to *everything* she's learned about the evidence as she knows the evidence would, which includes both $[P^1(p) \geq .7]$ *and* $[P^1(p) \leq .7]$. As we've seen, upon learning both she should move to credence *.6*, not .7—that was our original Reflection failure. Thus trying to "recover" Reflection by repeatedly applying Trust would require ignoring information you receive along the way.

Upshot: The failure of Reflection motivates New Reflection, who's slogan and failure in turn motivate Trust. Trust, then, is a formally elegant and philosophically well-founded principle connecting first- and higher-order evidence.

Perhaps, though, you've become skeptical of the prospects of such principles (cf. Lasonen-Aarnio, 2015). Fear not: here come the fireworks.

5.1 Trust's Rewards

We have subsumption results. Trust provably generalizes New Reflection:

Proposition* 5.1. *Trust implies New Reflection, but not vice versa.*[13]

We have efficacy results. Trust banishes our puzzles:

Proposition* 5.2. *Trust is inconsistent with Improbable Knowing, Misguided Evidence, and Self-Effacing Evidence.*

This is for exactly the reason you'd expect. By forcing a connection between evidence and truth, Trust prevents you from expecting the evidence to misguide you. But this is only half the battle. It's *easy* to rule out puzzles; much harder to do so without trivializing higher-order uncertainty. Here Trust comes into its own.

We have tenability results. Trust allows plenty of higher-order uncertainty:

Proposition* 5.3. *For any $\epsilon > 0$ there are Trust-validating frames for which $\exists p, w : \forall t : P^i_w(P^i(p) = t) < \epsilon$.*

That is, Trust allows you to have no idea what your evidence supports.

So Trust yields the goods we're after. But we want to know more: what sort of picture of evidence and rational belief does it offer? For that, turn to the model theory. We have characterization results. Trust axiomatizes an elegant and under-explored class of Kripke frames composed of structures like Figure 2. Dots are worlds; circles are drawn around worlds that see exactly the same worlds; transitive arrows are omitted. Formally, such R^i are transitive, surely-reflexive, and surely-nested—definitions are in Appendix B (Theorem 5.4).

What do you see? A *tree!*[14] Such structures are formally tractable and mathematically elegant—if this is the structure of evidence, mathematicians will be pleased.

I claim that it is:

Theorem 5.4 (Characterization)**.** *A probabilistic frame $\langle W, R^1, R^2, \mu\rangle$ validates Trust iff $\langle W, R^1, R^2\rangle$ is transitive, surely-reflexive, surely-updating, and surely-nested.*

[13] A '*' indicates that the result holds up under any model theory—see the end of Appendix B.

[14] Formally, if we were to coarsen the frame by taking equivalence classes under neighborhoods and force the accessibility relation asymmetric, the resulting structure of every neighborhood would become a *forest* in the ideology of graph theory. Dubbed **neighborhood forests**, since every neighborhood becomes a forest under this neighbordhood-coarsening.

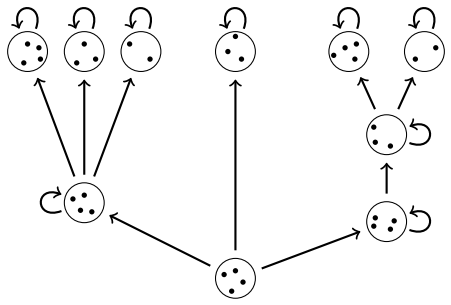


Figure 2: A Neighborhood Forest

Moreover, such structures admit of a natural interpretation—philosophers will be pleased as well.

Begin with an example: I hand the engineers a file of evidence, asking whether the plane is safe. What would be a reasonable way for them to carry out the inquiry? Settle all the (relevant) questions they can, and then estimate the likelihood that the plane is safe, given their answers. They should start with the easy ones: "What model is the plane? What statistics do we have on those?..." Move to the more difficult: "How many landings can the average wheel sustain? How much friction do these turbines generate?..." Eventually they'll arrive at questions they're not in a position to answer—as the case may be, perhaps, "Do the controls handle smoothly?" At this point the rational thing to do is to stop settling questions and simply estimate likelihoods based on their answers so far. Of course, whether they're in a position to settle a given question can *itself* be a difficult question—sometimes they should be unsure exactly where to stop, as Mil is in our original case.

This picture should feel familiar—think of flow charts and decision-trees, wherein you answer questions in stages, with your answers opening up yet new questions. With that in mind, take a look back at Figure 2. What do you see? **Inquiry as question-settling.** Think of a given inquiry as taking place within a space of relevant questions. At each stage of information-processing (rational belief-formation) you *settle a question*—in the frame, you proceed down one of the branch-points. Your answer then becomes a fixed point in later processing, affecting the relevant questions, the available answers, and the relative likelihoods in what follows. There are some questions you should be sure you *can't* settle—in the frame, those are the questions left unsettled in the "leaves" (top nodes) of our tree. There are other questions you should be sure you *can* settle—those are the branch-points "behind" you in the tree. But there are other questions you should be unsure about—maybe you can settle them, maybe you can't. These are the branch-points "ahead" of you in the tree—questions that perhaps your evidence *does* settle. Therein lies your higher-order uncertainty.

This interpretation is only a sketch—I explore it elsewhere. But I think it's a very natural sketch of the way we do and should use available evidence. To illustrate: we have applicability results. We can now construct a systematic, well-behaved model of The Unmarked Clock.

Suppose that when you glimpse the clock there are two stages of information processing. First you figure out your *best guess* for the hand's position; then you figure out what your *margin for error* is, given that guess. From there, you estimate likelihoods of various positions. Formalizing, we have Figure 3:[15]

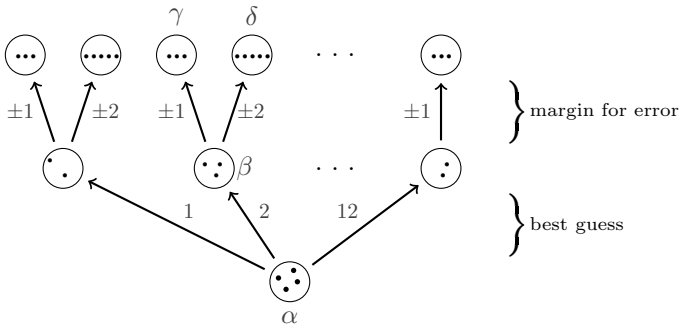


Figure 3: The Trustworthy Clock

This is a schematic model—we could fill in the worlds and probabilities in different ways. α includes radical skeptical scenarios where you don't know what your best guess is. γ is where you should settle both that your guess is 2 and that your margin is ± 1, so you should be sure it's pointing at 1,2, or 3—and, importantly, you should be sure *that* you should be sure of this. δ is similar except you should settle that your margin is ± 2. β is where you should be sure your guess is 2, but unsure whether you can figure out what your margin for error is—maybe you should settle that it's ± 1, maybe that it's ± 2, maybe neither. And so on.

There is much to discuss about the details of the model, but the crucial points are these. (1) The "branching" structure of this model is motivated from a systematic background theory—*Trust the evidence*—not an ad hoc patch. (2) Such structure forces us to remove the fiction that the hand's *actual* position fully determines your epistemic state—which is what led the probability of *odd* to be anti-correlated with its truth. (3) Unlike other reconstructions of cases like The Unmarked Clock, ours allows higher-order uncertainty by allowing you to be unsure what your margin for error (or best guess) is (cf. Stalnaker, 2009; Hawthorne and Magidor, 2010; Cohen and Comesaña, 2013). (4)

[15]Dots represent worlds; arrows and circles represent R^2-relations (R^1 is universal—before your glimpse you don't know anything about the clock), with reflexive and transitive arrows omitted; everything else is labelling.

Finally, since Figure 3 describes a surely-updating neighborhood forest, by Theorem 5.4 it validates Trust, and so by Proposition 5.2 it *in*validates Improbable Knowing, Misguided Evidence, and Self-Effacing Evidence.

Upshot: Trust is a well-motivated, non-trivializing Bridging principle that provides a systematic way to model cases while avoiding puzzles of higher-order evidence. Bridging succeeds, after all.

6 *The* Theory

But though we have *a* solution, we so far have no argument that it's *the* solution. What if Trust is too weak—permitting as-yet-unnamed puzzles? What if Trust is too strong—ruling our more than is required to solve them?

We can offer a *proof*, of sorts, that it's not: no stronger theory is needed, and no weaker theory will do. How? We'll independently characterize what it takes for evidence to be a guide, and then show that this characterization leads *exactly* to our theory. Exactly to Trust.

Why should we treat the evidence as a guide? Because it's *valuable*: we should expect it to put us in a better position to fulfill our ends. Example: Mil is faced with the decision of whether to send the plane to inspection. At worlds where it's safe, this is good; at those were it's not, it's bad. If she were .9 confident it's safe, she'd send; if she were .5, she'd hold off. Being .7, she's on the fence—the expected value of sending and holding off is balanced. What about the expected value of doing what she *should* do—whatever that is? If she should be .9 (as perhaps she should), what she should do is send; if she should be .5 (as perhaps she should), what she should do is hold off. Being unsure what she should think, she's unsure what she should do. But why care about doing what she *should* do? Because she should *value* the evidence—expect that doing as it recommends will lead to a better outcome than otherwise. That's why she should pour over the evidence to figure out what she *should* think and do, rather than just going with her gut—because she should expect that if she succeeds, she'll make a better decision.

A version of this idea was made famous by I.J. Good (1967). Suppose you face a *decision problem* $\langle O, U\rangle$ modeled by a set of options O and a real-valued utility function U. For each option $o \in O$, $U(o, w)$ is the value of performing option o at world w. Then the evidence is *valuable* iff—were it cost free[16]—you should prefer to make use of it to

[16]Of course, there's no such thing as a free lunch—information is never free, and it's often not worth the (monetary or psychological) cost. Likewise, there's no such thing as a frictionless plane. But just as the physicist sets aside friction to measure the true force of gravity, so the epistemologist sets aside psychology to measure the true value of evidence.

guide your decision: iff the expected value of doing so is higher than that of simply choosing an option.

In formulating this question, Good and his followers have focused on the *diachronic* setting: Should Mil prefer to use a *more* informed body of evidence to make her decision?[17] But once we recognize the rationality of higher-order uncertainty, there's an equally pressing question: Should Mil prefer to use her *current* evidence—whatever it is—to make her decision? If evidence is to play the guiding role we're after, then the answer to both questions must be "yes." The expected value of using a body of evidence *at least as informed as your own* must be higher than simply choosing an option.

We can formalize this in probabilistic epistemic logic. $[U(o) = s]$ is the proposition that option o yields value s: $\boldsymbol{[U(o) = s] =_{df} \{w|U(o, w) = s\}}$. The *expected* value of o is an average of the various possible values it might take, with weights determined by how confident you should be in each. Formalizing: at w the expected value of o given evidence i is $\boldsymbol{E^i_w[U(o)] =_{df} \sum_s P^i(U(o) = s) \cdot s}$.[18]

What about the expected value (given evidence i) of doing as you *should* do?—Of using evidence k ($k \geq i$) to make your decision? Since evidence k varies across worlds, what you should *do* with it varies across worlds—if you're at w, what you should do is take an option o that maximizes expected value by the lights of P^k_w. So just as we have a function P^k from worlds to probabilities that captures what you should *think* given evidence k, so too we'll have a function D^k from worlds to options that captures what you should *D*o given evidence k. Formalizing: D^k is a function from worlds w to options $D^k_w \in O$ such that $E^k_w[U(D^k_w)] = \max_{o \in O} E^k_w[U(o)]$.[19] This allows us to define the proposition $[U(D^k) = s]$ that the value of doing what you *should* do given evidence k is s: $\boldsymbol{[U(D^k) = s] =_{df} \{w|U(D^k_w, w) = s\}}$. Then at w the expected value (given evidence i) of letting evidence k guide your decision is $\boldsymbol{E^i_w[U(D^k)] =_{df} \sum_s P^i_w(U(D^k) = s) \cdot s}$.

We can now state our constraint. You should *value* a body of evidence iff you should expect that letting it guide your decision will make you better off than any other particular option o. Formalizing:

$$\textbf{Value:} \quad E^i_w[U(D^k)] \geq E^i_w[U(o)] \qquad (k \geq i)$$

[17] E.g. Good (1967); Skyrms (1990); Huttegger (2014); Myrvold (2012).

[18] Decision theorists will notice that here I've used Savage (1954) rather than causal or evidential decision theory—meaning I've assumed that probabilities are independent of which option is in question. I do so because it provides the purest metric of the intrinsic value of the information, unaffected by whether taking an option yields *new* information. It is a messy and difficult question whether the value of evidence does (or should) hold when we allow *which option* you take to affect your evidence. But for my purposes the point is that it'd *better* hold in this case.

[19] Since D^1 and D^2 encode ways of responding to bodies of evidence, we impose the constraint that if you have the same information at worlds x and y, then what you should do with it is the same as well: if $P^i_x = P^k_y$ then $D^i_x = D^k_y$.

Claim: Value is a formal statement of the value of evidence.[20] Our first truism is true—you should treat the evidence as a guide—iff the correct theory of evidence validates Value.[21]

What does it take to do so? According to legend, Good (1967) proved the value of evidence as a theorem of decision theory. But that can't be right, for we've already seen counterexamples—to wit, our puzzles of higher-order evidence.

Proposition* 6.1. *Value is inconsistent with Improbable Knowing, Misguided Evidence, and Self-Effacing Evidence.*

So Value is no theorem of decision theory. What of Good's proof? He implicitly assumes that the accessibility relations R^i are partitional. That is to trivialize higher-order uncertainty—any partitional probabilistic frame validates $[P^i(p) = t] \leftrightarrow S^i[P^i(p) = t]$. Good's proof will not avail us. For recall our question: How can evidence be a guide, *given* the rationality of higher-order uncertainty?

Answer: *Trust the evidence.* We have coincidence results:

Theorem 6.2 (Value of Evidence Theorem)**.** *The following are equivalent:*

(1) *The probabilistic frame* $\langle W, R^1, R^2, \mu\rangle$ *validates Trust.*

(2) $\langle W, R^1, R^2\rangle$ *is transitive, surely-reflexive, surely-updating, and surely-nested.*

(3) *The probabilistic frame* $\langle W, R^1, R^2, \mu\rangle$ *validates Value.*

You should treat the evidence as a guide *if and only if* you should trust the evidence.

I call this a "coincidence result," for that is exactly what it is. The progression of the project was not so prescient as the progression of this paper. I began with puzzles of higher-order evidence, was led (through trial and error) to Trust, and characterized it over the class of transitive, surely-reflexive, surely-updating, and surely-nested frames. Only later did I notice a strikingly similar result:

Theorem 6.3 (Geanakoplos)**.** *A frame* $\langle W, R^1, R^2\rangle$ *with* $R^1 = W \times W$ *validates Value* under every prior μ *iff* $\langle W, R^1, R^2\rangle$ *is transitive, reflexive, and nested.*[22]

Imagine my surprise—shock, even—upon seeing this theorem. And my satisfaction upon discovering that it could be strengthened, yielding our perfect little coincidence. It is a significant fact that we end up at the exact same destination from two—very

[20] An important special case: letting O be the set of credences and U be an *accuracy* metric à la epistemic utility theory (Joyce, 1998; Pettigrew, 2016), Value implies that you should expect the evidence to make your beliefs more accurate.

[21] A probabilistic frame $\langle W, R^1, R^2, \mu\rangle$ validates Value iff for every decision problem $\langle O, U\rangle$—no matter what options and values you have—$\langle W, R^1, R^2, \mu, O, U\rangle$ validates Value.

[22] Geanokoplos is using a slightly different framework, so this is the closest, easily-statable version of his theorem in our setup. Many thanks to [XXX], who pointed me to Geanakoplos (1989) long before any of this had been worked out—his skeptical challenges have paid off in leaps and bounds.

different, very well-motivated—starting points. Such coincidences do not happen—in mathematics or in philosophy—unless that destination is a place worth going.

Upshot: Trust is an epistemic characterization of the value of evidence. No stronger theory is needed, and no weaker theory will do.

7 Conclusion

We began with the problem of higher order evidence: How can evidence be a guide, given the rationality of higher-order uncertainty? We've now found a—*the*—solution: *Trust the evidence.*

Trust will no doubt face objections—to its consequences, its presuppositions, its idealization. But theories—whether epistemological, mathematical, or scientific—are to be judged by their fruits. Ours can claim the following. It is completely general, formally precise, and philosophically versatile. It formalizes a compelling idea. It refines the insights from previous promising approaches. It vindicates the rationality and import of higher-order uncertainty, while unifying and banishing persistent puzzles of it. It has an elegant and tractable mathematical structure. It can be given a natural and systematic interpretation. It offers a new picture of rational belief-formation. It guarantees that evidence is a guide to truth. And it vindicates and characterizes the value of evidence.

Finally, assume the E=K thesis that evidence is knowledge (Williamson, 2000). Then Trust—hence the value of evidence—implies the KK principle: that if you know something, you're in a position to know that you do. We have a grand argument for KK.

References

Christensen, D. (2010a). Higher-order evidence. *Philosophy and Phenomenological Research*, 81(1):185–215.

Christensen, D. (2010b). Rational reflection. *Philosophical Perspectives*, 24:121–140.

Christensen, D. (2016). Disagreement, drugs, etc.: From accuracy to akrasia. *Episteme*, Forthcoming.

Coates, A. (2012). Rational epistemic akrasia. *American Philosophical Quarterly*, 49(2):113–124.

Cohen, S. and Comesaña, J. (2013). Williamson on gettier caes in epistemic logic. *Inquiry*, 56(1):15–29.

Das, N. (ms). Fragmented evidence.

Elga, A. (2007). Reflection and disagreement. *Noûs*, 41(3):478–502.

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

Feldman, R. (2005). Respecting the evidence. *Philosophical Perspectives*, 19(1):95–119.

Gaifman, H. (1988). A theory of higher order probabilities. In Skyrms, B. and Harper, W. L., editors, *Causation, Chance, and Credence*, volume 1, pages 191–219. Kluwer.

Geanakoplos, J. (1989). Game theory without partitions, and applications to speculation and consensus. Cowles Foundation Discussion Paper 914, Yale University.

Gibbons, J. (2006). Access externalism. *Mind*, 115(457):19–39.

Good, I. J. (1967). On the principle of total evidence. *The British Journal for the Philosophy of Science*, 17(4):319–321.

Greco, D. (2014). A puzzle about epistemic akrasia. *Philosophical Studies*, 161:201–219.

Hawthorne, J. and Magidor, O. (2010). Assertion and epistemic opacity. *Mind*, 119(476):1087–1105.

Hazlett, A. (2012). Higher-order epistemic attitudes and intellectual humility. *Episteme*, 9(3):205–223.

Hintikka, J. (1962). *Knowledge and Belief*. Cornell University Press.

Horowitz, S. (2014). Epistemic akrasia. *Noûs*, 48(4):718–744.

Huemer, M. (2011). The puzzle of metacoherence. *Philosophy and Phenomenological Research*, 82(1):1–21.

Huttegger, S. M. (2014). Learning experiences and the value of knowledge. *Philosophical Studies*, 171(2):279–288.

Joyce, J. M. (1998). A non-pragmatic vindication of probabilism. *Philosophy of Science*, 65(4):575–603.

Lasonen-Aarnio, M. (2010). Unreasonable knowledge. *Philosophical Perspectives*, 24(1):1–21.

Lasonen-Aarnio, M. (2014). Higher-order evidence and the limits of defeat. *Philosophy and Phenomenological Research*, 8(2):314–345.

Lasonen-Aarnio, M. (2015). New rational reflection and internalism about rationality. In Gendler, T. S. and Hawthorne, J., editors, *Oxford Studies in Epistemology*, volume 5, pages 145–171. Oxford University Press.

Littlejohn, C. (2016). Stop making sense? on a puzzle about rationality. *Philosophy and Phenomenological Research*, Forthcoming.

Myrvold, W. C. (2012). Epistemic values and the value of learning. *Synthese*, 187(2):547–568.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press.

Rasmussen, M. S., Steglich-Petersen, A., and Bjerring, J. C. (forthcoming). A higher-order approach to disagreement. *Episteme*.

Roush, S. (2016). Knowledge of our own beliefs. *Philosophy and Phenomenological Research*, pages n/a–n/a.

Salow, B. (2016). The externalist's guide to fishing for compliments. *Mind*, Forthcoming.

Salow, B. (ms). Elusive externalism.

Samet, D. (1997). On the triviality of high-order beliefs. page https://ideas.repec.org/p/wpa/wuwpga/9705001.html.

Savage, L. J. (1954). *The Foundations of Statistics*. Wiley Publications in Statistics.

Schoenfield, M. (2015). A dilemma for calibrationism. *Philosophy and Phenomenological Research*, 91(2):425–455.

Schoenfield, M. (2016). An accuracy based approach to higher order evidence. *Philosophy and Phenomenological Research*, Forthcoming.

Skyrms, B. (1980). Higher order degrees of belief. In Mellor, D. H., editor, *Prospects for Pragmatism*, pages 109–137. Cambridge University Press.

Skyrms, B. (1990). The value of knowledge. *Minnesota Studies in the Philosophy of Science*, 14:245–266.

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

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

Smithies, D. (2015). Ideal rationality and logical omniscience. *Synthese*, 192(9):2769–2793.

Smithies, D. (ms). The irrationality of epistemic akrasia.

Stalnaker, R. (2009). On hawthorne and magidor on assertion, context, and epistemic accessibility. *Mind*, 118(470):39–49.

Titelbaum, M. (2015). Rationality's fixed point (or: In defense of right reason). In Gendler, T. S. and Hawthorne, J., editors, *Oxford Studies in Epistemology*, volume 5, pages 253–292. Oxford University Press.

van Ditmarsch, H., Halpern, J. Y., van der Hoek, W., and Kooi, B. (2015). *Handbook of Epistemic Logic*. College Publications.

van Fraasen, B. (1984). Belief and the will. *The Journal of Philosophy*, 81(5):235–256.

Vavova, K. (2014). Confidence, evidence, and disagreement. *Erkenntnis*, 79:173–183.

Vavova, K. (2016). Irrelevant influences. *Philosophy and Phenomenological Research*, Forthcoming.

Weatherson, B. (ms). Do judgments screen evidence?

Wedgwood, R. (2012). Justified inference. *Synthese*, 189:273–295.

White, R. (2009). On treating oneself and others as thermometers. *Episteme*, 6(3):233–250.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.

Williamson, T. (2014). Very improbable knowing. *Erkenntnis*, 79(5):971–999.

Worsnip, A. (2016). The conflict of evidence and coherence. *Philosophy and Phenomenological Research*, Forthcoming.

Appendix A: Probabilistic Epistemic Logic

We'll build a probabilistic frame $\langle W, R^1, R^2, \mu\rangle$ to model our case of The Engineers (§3) in stages.

Start with the possibilities. $\boldsymbol{W}$ is a (finite) set of worlds, thought of as a partition that captures the relevant distinctions for modeling the case at hand. Our case has three such distinctions: (1) Is the plane safe, or not? (2) Are the controls smooth, or not? (3) Does the evidence settle whether the controls are smooth, or not? Mixing and

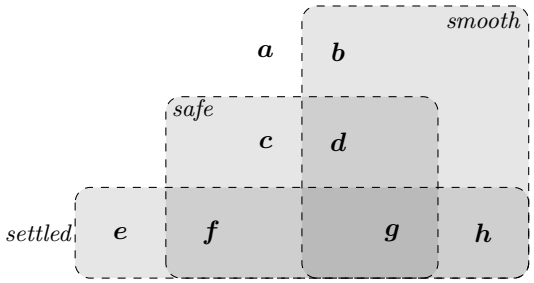


Figure 4: The Engineers $\langle W \rangle$

matching, we have 8 possibilities: $W = \{a, b, c, d, e, f, g, h\}$. Our propositions can then be represented in a Venn diagram as sets of worlds, as in Figure 4.
Shaded regions represent propositions, e.g. $safe = \{c, d, f, g\}$. Instead of using an official object language, we'll use propositional functions to handle logical operations: if p and q are propositions (subsets of W), $\neg p$ is p's complement $W - p$; $p \wedge q$ is their intersection $p \cap q$, etc. p is true at a world w iff $w \in p$ and p entails q just in case every p-possibility is a q-possibility, i.e. $p \subseteq q$. So (e.g.) $smooth \wedge settled$ is true at h, since $h \in \{b, d, g, h\} \cap \{e, f, g, h\} = \{g, h\}$; and $smooth$ implies $smooth \vee safe$, since $\{b, d, g, h\} \subseteq \{b, c, d, f, g, h\} = \{b, d, g, h\} \cup \{c, d, f, g\}$.

Next we want to model Mil's evidence today. $\boldsymbol{R^1}$ is a (serial) binary relation on W. xR^1y means at world x Mil's evidence today leaves open that she's at world y—we say "x accesses/sees y." We can represent Mil's evidence today by enriching our diagram to Figure 5.

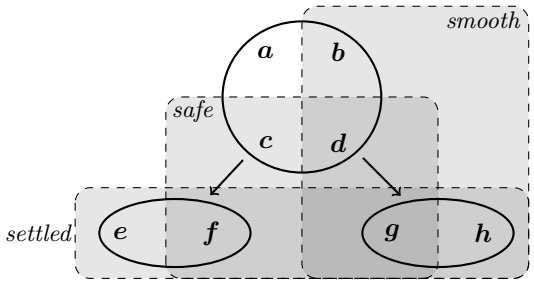


Figure 5: The Engineers $\langle W, R^1 \rangle$

Here ovals are drawn around worlds that see exactly the same worlds (so a, b, c, d all see the same possibilities); black arrows represent R^1-relations, and an arrow pointing to an oval means all worlds inside it are seen (so a, b, c, d sees g and h, but g and h don't see them). The **(1-)neighborhood** of world w is R^1_w—the set of possibilities consistent with Mil's evidence (today) at world w: $\boldsymbol{R^1_w =_{df} \{x | wR^1x\}}$.

Thus $R^1_h = \{g, h\}$ while $R^1_a = \{a, b, c, d, e, f, g, h\}$. It varies across worlds because Mil's evidence does. Moreover, the fact that a sees both a and h but $R^1_a \neq R^1_h$ means that Mil has higher-order uncertainty: at a she should leave open both that he evidence is R^1_a (i.e. it can't rule out any possibilities in W) and that it's R^1_h (i.e. that it settles that the controls are smooth).

Precisely: we can use R^1 to define propositions about Mil's evidence today. If p is a proposition, S^1p is the proposition that she should be *S*ure of p given evidence 1—that p is *S*ettled by this evidence. It's true at w iff every world consistent with her evidence at w is a p-world: $\boldsymbol{S^1p =_{df} \{w | R^1_w \subseteq p\}}$.[23] Thus letting $p = smooth = \{b, d, g, h\}$, $S^i\neg p = \{e, f\}$, $S^ip = \{g, h\}$, and $\neg S^ip \wedge \neg S^i\neg p = \{a, b, c, d\}$. In words: Lois is rational at $\{e, f\}$ (where their evidence settles that the controls are not smooth), Hiedi is rational at $\{g, h\}$ (where their evidence settles that they *are* smooth), and Mil is rational at $\{a, b, c, d\}$ (where their evidence doesn't settle either way). And since $\{a, b, c, d\}$ leaves open $\{e, f\}$ and $\{g, h\}$, if Mil's rational then she can't rule out that Lois or Hiedi is: $\neg S^i\neg(S^i\neg p)$ and $\neg S^i\neg(S^ip)$ are true at $\{a, b, c, d\}$.

Next, we want to model Mil's evidence *tomorrow*, after they ask Eve the expert who responded to their evidence rationally. $\boldsymbol{R^2}$ is another (serial) binary relation, with parallel definitions for R^2_w, S^2p, etc. Enriching our diagram to let blue ovals/arrows represent R^2 relations, we get Figure 6.

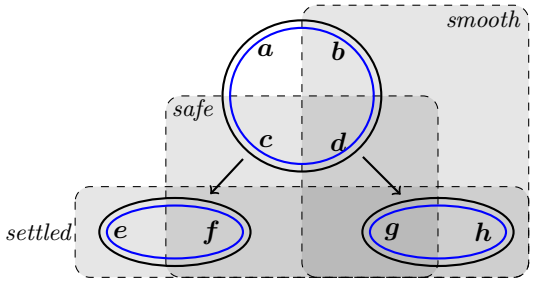


Figure 6: The Engineers $\langle W, R^1, R^2\rangle$

The difference between R^1 and R^2 is that there are no blue arrows from $\{a, b, c, d\}$ to any world that sees different worlds. That is, if Mil's rational then although *today* she should be uncertain of this, *tomorrow*—after Eve tells her—she shouldn't be.

So far this is standard epistemic logic. But we want to add *degrees* of evidential support—*probabilistic* epistemic logic.[24] We accomplish this by modeling Mil's (ra-

[23] Since R^i is not assumed to be reflexive, S^i is not assumed to be factive: $S^ip \rightarrow p$ is not automatically valid. That means S^i needn't be interpreted as knowledge—we needn't be knowledge-first epistemologists to use epistemic logic.

[24] My approach here is most similar to that of Williamson (2000, 2014), though he offers a slightly different interpretation. Similar formalisms are commonly used in the literature on epistemic logic—

tional) background standards of reasoning: $\boldsymbol{\mu}$ is a (regular) probability distribution over W. It captures how likely Mil should think each possibility is, *absent the evidence in question*—ignoring R^1 and R^2. This gives us Figure 7:

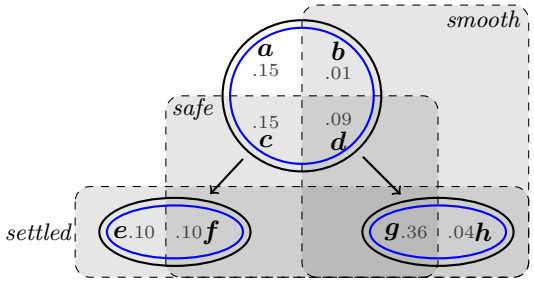


Figure 7: The Engineers $\langle W, R^1, R^2, \mu \rangle$

Here is our final probabilistic frame. The background probability of a world is the number next to it, and that of a *proposition* is the sum of the probabilities of its worlds, e.g. $\mu(safe) = \mu(\{c, d, f, g\}) = \mu(c) + \mu(d) + \mu(f) + \mu(g) = .7$.

If the controls are smooth, it's .9 likely the plane is safe: $\mu(safe|smooth) = .9$. If not, it's .5 likely: $\mu(safe|\neg smooth) = .5$. Moreover, recall that Mil both (1) is slightly inclined to think the controls are *not* smooth, but (2) thinks its more likely that Hiedi's right than that Lois is. Thus (1) $\mu(smooth|\{a, b, c, d\}) = .25$, but (2) $\mu(\{g, h\}) = .4 > .2 = \mu(\{e, f\})$. These are Mil's background standards of reasoning.

We obtain what she should think given her *total* evidence (today) at a world w, written P^1_w, by conditionalizing her standards μ on that evidence R^1_w to get $\boldsymbol{P^1_w(p) =_{df} \mu(p|R^1_w) = \frac{\mu(p \cap R^1_w)}{\mu(R^1_w)}}$. Similarly for what she should think tomorrow: P^2_w is μ conditionalized on R^2_w. Since Mil's evidence varies across worlds and times, what she should think does too: $P^1_a(safe) = .7$, but $P^1_{\boldsymbol{h}}(safe) = .9$, $P^{\boldsymbol{2}}_{\boldsymbol{a}}(safe) = .6$, and $P^{\boldsymbol{2}}_{\boldsymbol{h}}(safe) = .9$. Thus today at world a Mil should be uncertain both what her evidence now supports and what her evidence tomorrow will support.

Precisely: just as we used R^i_w to define propositions about what Mil should be sure of, we can use P^i_w to define propositions about how confident she should be. For each proposition p and $t \in [0, 1]$, define $[P^1(p) = t]$ to be the set of worlds where Mil's evidence (today) makes p exactly t-likely: $\boldsymbol{[P^1(p) = t] =_{df} \{w|P^1_w(p) = t\}}$, and similarly for $[P^2(p) = t]$ (and other facts about probabilities, like conditional ones: $[P^i_q(p) = t] = [P^i(p|q) = t] =_{df} \{w|$ if $P^i_w(q) > 0$ then $P^i_w(p|q) = t\}$.) Thus $[P^1(safe) = .5] = \{e, f\}$, $[P^1(safe) = .9] = \{g, h\}$, and $[P^1(safe) = .7] = \{a, b, c, d\}$. In words: Lois's .5 credence is rational at $\{e, f, \}$, Hiedi's .9 credence is rational at $\{g, h\}$, and Mil's .7 credence is

though usually with the (in our case, trivializing) assumption that agents are certain of their own probabilities. See e.g. (van Ditmarsch et al., 2015, Ch. 4).

rational at $\{a, b, c, d\}$.

More: since at $\{a, b, c, d\}$ both $[P^1(\{e, f\}) = .2]$ and $[P^1(\{g, h\}) = .4]$, Mil should have .2 credence that Lois is right and .4 credence that Hiedi is. That means that facts about higher-order probabilities fall right out of the model. If Mil's rational, she should have .2 credence that she should have .5 credence the plane's safe, and .4 credence that she should have .9 credence it's safe: at $\{a, b, c, d\}$. both $[P^1(P^1(safe) = .5) = .2]$ (since $[P^1(safe) = .5] = \{e, f\}$) and $[P^1(P^1(safe) = .9) = .4]$ (since $[P^1(safe) = .9] = \{g, h\}$). Moreover, if she's rational she should also be uncertain today what her evidence *tomorrow* supports: at $\{a, b, c, d\}$, $[P^1(P^2(safe) = .9) = .4]$, for instance. In contrast, tomorrow after they talk to Eve, Mil should have *no* higher-order uncertainty—for instance, at $\{a, b, c, d\}$ both $[P^2(safe) = .6]$ and $[P^2(P^2(safe) = .6) = 1]$.

These are the models of higher-order uncertainty that I will use to build and test our theories, following the tradition of Hintikka (1962). We *could* be more general. Instead of generating probabilities from a prior μ we could simply use a function $\mathcal{P}^i$ from worlds to probability functions $\mathcal{P}^i_w$, which could vary unconstrained. If we did, we'd be using **general probabilistic frames** $\langle W, \mathcal{P}^1, \mathcal{P}^2\rangle$—the most general models needed to study higher-order probability. Alas, their flexibility makes them much less formally tractable than probabilistic frames. But *many* of our results hold up under this model theory—they are marked with a '*'. I conjecture that all our results can be so generalized, but that is a big formal project.

Appendix B: Proofs

Propositions 5.1-6.3

Proposition 4.1. *New Reflection is validated by The Unmarked Clock.*

Proof. Take an arbitrary world w and instance of the probability in New Reflection $P^i_w(p|P^k = \pi)$. We show this equals $\pi(p|P^k = \pi)$. Suppose $k = 1$ (hence $i = 1$). Since R^1 is trivial, $[P^1 = \mu] = W$ so for $\pi \neq \mu$ this is undefined (so holds trivially) and for $\pi = \mu$, $P^1_w(p|P^1 = \mu) = P^1_w(p) = \mu(p) = \mu(p|P^1 = \mu)$, as desired. Suppose $k = 2$. Note that each world x has a unique 2-neighborhood R^2_x, implying that it has a unique probability function P^2_x. The only nontrivial instances of New Reflection are when $\pi = P^2_x$ for some such x. And, since unique, $[P^2 = P^2_x] = \{y|P^2_y = P^2_x\} = \{x\}$ for each x. So consider $P^i_w(p|P^2 = P^2_x)$, and suppose its well-defined. Then it equals (a) $P^i_w(p|x)$. Since this frame is reflexive, $P^2_x(p|P^2 = P^2_x)$ is well-defined and equal to (b) $P^2_x(p|x)$. If $x \in p$, then (a) = (b) = 1; and if $x \notin p$, then (a) = (b) = 0. Thus $P^i_w(p|P^2 = P^2_x) = P^2_x(p|P^2 = P^2_x)$, as desired. □

Proposition* 5.1. *Trust implies New Reflection, but not vice versa.*

Proof. By Proposition 4.1 New Reflection is consistent with Improbable Knowing, but by Proposition 5.2 Trust is not; so New Reflection does not imply Trust.

On the other hand, suppose New Reflection is false at w: $\exists \pi, p : P^i_w(p|P^k = \pi) \neq \pi(p|P^k = \pi)$. Without loss of generality, suppose $P^i_w(p|P^k = \pi) < \pi(p|P^k = \pi)$. We will show that Trust fails at w. Define l, h such that $l = P^i_w(p|P^k = \pi) < \pi(p|P^k = \pi) = h$. We first show that $\boldsymbol{(\alpha)}$ **:** $[P^k = \pi] = \big([P^k = \pi] \wedge [P^k(p|P^k = \pi) \geq h]\big)$. For take any $y \in [P^k = \pi]$. Since $P^k_y = \pi$, $P^k_y(p|P^k = \pi) = \pi(p|P^k = \pi) = h$. So $[P^k(p|P^k = \pi) \geq h]$ is true at y. y was an arbitrary member of $[P^k = \pi]$, so $[P^k = \pi] \subseteq [P^k(p|P^k = \pi) \geq h]$, implying that $\big([P^k = \pi] \wedge [P^k(p|P^k = \pi) \geq h]\big) = [P^k = \pi]$, as desired. Now, we know $P^i_w(p|P^k = \pi) = l < h$. By (α), we can substitute to get $\boldsymbol{(\beta)}$ **:** $P^i_w\big(p\big|[P^k = \pi] \wedge [P^k(p|P^k = \pi) \geq h]\big) < h$. Yet putting $p = p$, $q = [P^k = \pi]$, and $t = h$, an instance of Trust is $P^i_{[P^k=\pi]}\big(p\big|P^k_{[P^k=\pi]}(p) \geq h]\big) \geq h$, i.e. $P^i\big(p\big|[P^k = \pi] \wedge [P^k(p|P^k = \pi) \geq h]\big) \geq h$—which, by (β), is false at w. Thus if New Reflection is false at w, so is Trust. Contraposing, we have our result. □

Proposition* 5.2. *Trust is inconsistent with Improbable Knowing, Misguided Evidence, and Self-Effacing Evidence.*

Proof. Note: in finite, regular frames, $S^i p \leftrightarrow [P^i(p) = 1]$ is valid.

Improbable Knowing: Suppose $S^i p \wedge [P^i(S^i p) < \frac{1}{2}]$ is true at w. Since $S^i p$ is true, $P^i_w(p) = 1$; so $\boldsymbol{(\alpha)}$ **:** $P^i_w(p|q) = 1$ for any q on which it's defined. Since $P^i_w(S^i p) < \frac{1}{2}$, $P^i_w(\neg S^i p) > \frac{1}{2}$, i.e. $P^i_w\big(P^i(p) < 1\big) > \frac{1}{2}$. Thus $P^i_w\big(p\big|P^i(p) < 1\big)$ is well-defined, so by (α) $P^i_w\big(p\big|P^i(p) < 1\big) = 1$. But an instance of Trust with $t = 1 - \epsilon$ yields $P^i\big(p\big|P^i(p) < 1\big) < 1$, so Trust fails at w.

Misguided Evidence: Suppose $P^i_w\big(p \wedge [P^i(p) < t]\big) \geq t$. Since $P^i_w\big(P^i(p) < t]\big) \leq 1$, it follows that $\frac{P^i_w(p \wedge [P^i(p) < t])}{P^i_w(P^i(p) < t)} \geq t$. But an instance of Trust at $t - \epsilon$ yields $P^i\big(p\big|P^i(p) < t\big) < t$, which by the ratio formula implies $\frac{P^i(p \wedge [P^i(p) < t])}{P^i(P^i(p) < t)} < t$; so Trust fails at w.

Self-Effacing Evidence: Suppose $S^i(p \leftrightarrow [P^i(p) < \frac{1}{2}])$ is true at w. So for any $x \in [P^i(p) < \frac{1}{2}]$, if $P^i_w(x) > 0$ then $x \in p$. If there is such an x, $P^i_w(p|P^i(p) < \frac{1}{2}) = 1$, contradicting Trust. If not, then (by seriality) there must be a $y \in \neg[P^i(p) < \frac{1}{2}] = [P^i(\neg p) \geq \frac{1}{2}]$ such that $P^i_w(y) > 0$. Any such y must be a $\neg p$-world. Hence $P^i_w(p|P^i(p) \geq \frac{1}{2}) = 0$, contradicting Trust. □

Proposition* 5.3. *For any $\epsilon > 0$ there are Trust-validating frames for which $\exists p, w : \forall t : P^i_w(P^i(p) = t) < \epsilon$.*

Proof. By Theorem 5.4 it will suffice to construct a surely-updating neighborhood forest that meets the requirement. Let N be the smallest integer larger than $\frac{1}{\epsilon}$. Let $W = \{a_1, b_1, a_2, b_2, ..., a_{2N}, b_{2N}\}$. Let $R^1 = W \times W$, making the frame surely-updating. Define R^2 so that $R^2_{a_1} = R^2_{b_1} = W$, while for each $1 < i \leq 2N : R^2_{a_i} = R^2_{b_i} = \{a_i, b_i\}$. R^2

is a neighborhood forest: a_1 and b_1 see everything, while all other a_i and b_i see only themselves. Let $p = \{a_1, a_2, ..., a_{2N}\}$. Choose $2N$ different real numbers $u_i \in [0, 1]$ and set $\mu(a_i) = \frac{u_i}{2N}$ and $\mu(b_i) = \frac{1-u_i}{2N}$ for each i. Thus $\mu(\{a_i, b_i\}) = \frac{1}{2N}$, meaning $\mu(W) = 1$ as required. Notice that for each a_i, b_i with $i > 1$, $P^2_{a_i}(p) = P^2_{b_i}(p) = P^2_{b_i}(a_i) = u_i$. Since the u_i are unique, for $u_i \neq P^2_{a_1}(p)$ we have $[P^2(p) = u_i] = \{a_i, b_i\}$. Thus for $u_i \neq P^2_{a_i}(p)$, $P^2_{a_1}(P^2(p) = u_i) = \frac{1}{2N} < \epsilon$. And there is at most one other node $\{a_j, b_j\}, (j > 1)$ such that $P^2_{a_1}(p) = u_j$. Hence $P^2_{a_1}\big(P^2(p) = P^2_{a_1}(p)\big) \leq \frac{1}{N} < \epsilon$. Of course, for each other $t \in [0, 1]$, $P^2_{a_1}(P^2(p) = t) = 0 < \epsilon$. So at $w = a_1$ we have the desired result. □

Theorem 5.4: Characterizing Trust

Though the flagship result of this paper is the Value of Evidence Theorem (6.2), the largest original technical contribution is its first half: the characterization of Trust over (finite) probabilistic frames. Recall Trust: $P^i_q(p|P^k_q(p) \geq t]) \geq t$.

Some definitions. A frame is **surely-reflexive** iff every world that's seen by anything sees itself: $wR^ix \Rightarrow xR^kx$.[25] ("Surely" because every world is sure that the frame is reflexive!) A frame is **transitive** iff whenever x sees y and y sees z, x sees z: (xR^iy and yR^iz) $\Rightarrow$ xR^iz. Recall that the (i-)neighborhood of w is $R^i_w = \{x|wR^ix\}$—it includes all and only the worlds w can "see" under R^i; is it the strongest proposition you should be sure of at w given evidence i. A frame is **surely-updating** iff every world seen by anything has a smaller neighborhood (more information) under R^2 than R^1: $wR^ix \Rightarrow R^2_x \subseteq R^1_x$. A frame is **surely-nested** iff whenever anything sees x and y, if they can't see each other then they see nothing in common: if $x, y \in R^i_w$, then ($x\not{R}^ky$ and $y\not{R}^kx$) $\Rightarrow R^k_x \cap R^k_y = \emptyset$.

Given these definitions, we have:

Theorem 5.4 (Characterization)**.** *A probabilistic frame $\langle W, R^1, R^2, \mu\rangle$ validates Trust iff $\langle W, R^1, R^2\rangle$ is transitive, surely-reflexive, surely-updating, and surely-nested.*

A frame $\langle W, R^1, R^2\rangle$ is transitive, surely-reflexive, surely-updating, and surely-nested iff each of its neighborhoods R^i_w is transitive, reflexive, updating, and nested. Formally, iff $\forall x, y \in R^i_w$ we have the following. Transitive: $\forall z \in W : (zR^kx \wedge xR^ky) \Rightarrow zR^ky$ (equivalently: $x \in R^k_z \Rightarrow R^k_x \subseteq R^k_z$); reflexive: xR^kx; updating: $R^2_x \subseteq R^1_x$; nested: ($x\not{R}^ky$ and $y\not{R}^kx$) $\Rightarrow R^k_x \cap R^k_y = \emptyset$. (As always when indexing, $k \geq i$.) With this in hand, Theorem 5.4 breaks down into three lemmas:

Lemma 5.4.1. *If $\langle W, R^1, R^2, \mu\rangle$ validates Trust, it is transitive, surely-reflexive, surely-updating, and surely-nested.*

[25] Recall that we restrict indices so that $k \geq i$ in such contexts. Thus surely-reflexivity says three things: $wR^1x \Rightarrow wR^1x$, $wR^1x \Rightarrow wR^2x$, and $wR^2x \Rightarrow wR^2x$.

Lemma 5.4.2. *If R^i_w is transitive, reflexive, updating, and nested, so is any $q \subseteq R^i_w$.*

Lemma 5.4.3. *If R^i_w is transitive, reflexive, updating, and nested, then at w Naive Trust $\left(P^i(p|P^k(p) \geq t) \geq t\right)$ holds.*

We begin with Lemma 5.4.1:

Proof. We show the contrapositive. Take arbitrary $w \in W$; we show R^i_w satisfies:

Transitivity: Suppose $\exists x \in R^i_w$ such that $x \in R^k_z$ but $R^k_x \not\subseteq R^k_z$. By regularity, $[P^k(R^k_z) < 1]$ is true at x. Since $P^k_z(R^k_z) = 1$ and zR^kx, $P^k_z(R^k_z|P^k(R^k_z) < 1]) = 1$. Setting $q = W$ and $t = 1-\epsilon$, an instance of (downward) Trust is $P^k(R^k_z|P^k(R^k_z) < 1]) < 1$. Trust fails at z.

Reflexivity: Suppose $\exists x \in R^i_w$ such that $xR\!\!\!/^kx$. Let $p = W - \{x\}$, so $[P^k(p) = 1] \wedge \neg p$ is true at x. Since wR^ix, $P^i_w(\neg p|P^k(p) = 1) > 0$, so $P^i_w(p|P^k(p) \geq 1) < 1$; setting $q = W$ and $t = 1$, Trust fails.

Updating: We know R^i_w is reflexive. Suppose $\exists x \in R^i_w$ such that $R^2_x \not\subseteq R^1_x$, so there is a y with xR^2y but $xR\!\!\!/^1y$. By the latter, (α) $P^1_x(y) = 0$. By the former, $[P^2(y) > 0]$ is true at x. By reflexivity, xR^1x, so $P^1_x(P^2(y) > 0) > 0$. Combined with (α) we have $P^1_x(y|P^2(y) > 0) = 0$; setting $q = W$, $t = 1-\epsilon$, $i = 1$, and $k = 2$, Trust fails.

Nestedness: Suppose $\exists x, y, z \in R^i_w$ with $xR\!\!\!/^ky$ and $yR\!\!\!/^kx$ but $z \in R^k_x \cap R^k_y$. We know that R^i_w must be transitive, reflexive, and updating; we'll will show that Trust fails at w for $q = \{x, y, z\}$, $p = \{z\}$, and

$$t = \min_{v \in \{x,y,z\}} \left[P^k_v(p|q)\right].$$

By the definition of t, $[P^k(p|q) \geq t] \supseteq \{x, y, z\} = q \subseteq R^i_w$, so $\boldsymbol{(\alpha)}$ **:** $q = q \cap [P^k(p|q) \geq t] \cap R^i_w$. Now, $x, y \notin R^k_z$ for otherwise zR^kx or zR^ky, and so (by transitivity) xR^ky or yR^kx. Thus

$$P^k_z(p|q) = P^k_z(z|\{x, y, z\}) = 1. \qquad (\beta)$$

Moreover, $xR\!\!\!/^kw$ and $yR\!\!\!/^kw$, for otherwise (by reflexivity) wR^kw and (by transitivity) xR^ky or yR^kx. Hence:

$$P^k_x(p|q) = \frac{\mu(z \cap R^k_x \cap q)}{\mu(R^k_x \cap q)} = \frac{\mu(z)}{\mu(\{x, z\})} \qquad (\gamma)$$

$$P^k_y(p|q) = \frac{\mu(z \cap R^k_y \cap q)}{\mu(R^k_y \cap q)} = \frac{\mu(z)}{\mu(\{y, z\})} \qquad (\delta)$$

Combining (β), (γ), and (δ), we know

$$\begin{aligned} t &\geq \frac{\mu(z)}{\mu(\{x,z\})}, \frac{\mu(z)}{\mu(\{y,z\})} \\ &> \frac{\mu(z)}{\mu(\{x,y,z\})} && \text{(by regularity)} \\ &= \mu(p|q) \;=\; \mu(p|q \cap [P^k(p|q) \geq t] \cap R^i_w) && \text{(by } (\alpha)) \\ &= P^i_w(p|q \cap [P^k(p|q) \geq t]) \end{aligned}$$

That is, $P^i_q(p|P^k_q(p) \geq t) < t$ at w: Trust fails. □

To work further with updating forests, we first need to establish that they have the following fractal property:

Lemma 5.4.2. *If R^i_w is transitive, reflexive, updating, and nested, so is any $q \subseteq R^i_w$.*

Proof. Take arbitrary $x, y \in q \subseteq R^i_w$. Transitivity: say $x \in R^k_z$. Since $x \in R^i_w$ and R^i_w is transitive, $R^k_x \subseteq R^k_z$. Reflexivity: since $x \in R^i_w$ and R^i_w is reflexive, $x \in R^k_x$. Updating: since $x \in R^i_w$ and R^i_w is updating, $R^2_x \subseteq R^1_x$. Nested: Suppose $xR\!\!\!/^k y$ and $yR\!\!\!/^k x$. Since $x, y \in R^i_w$ and R^i_w is nested, $R^k_x \cap R^k_y = \emptyset$. Since x, y were arbitrary members of q, q is transitive, reflexive, updating, and nested. □

To prove the main step in the characterization, Lemma 5.4.3, we need some setup. Suppose we have a probabilistic frame $\langle W, R^1, R^2, \mu\rangle$ where each neighborhood R^i_w is transitive, reflexive, updating, and nested. Take an arbitrary such R^i_w.

Definition 5.4.3a (k-nodes)**.** Let the set $\boldsymbol{\mathcal{N}}$ of **k-nodes** partition R^i_w into worlds that see the same worlds under R^k: $\mathcal{N} =_{df} \{N \subseteq R^i_w | \forall x, y \in N : R^k_x = R^k_y\}$. (In our diagrams, k-nodes were the sets of worlds with circles around them.) The k-node of a world x is denoted $N_x = \{y \in R^i_w | R^k_y = R^k_x\}$. We let $\boldsymbol{A_w}$ denote the k-node whose members see *all* of R^i_w under R^k: $A_w =_{df} \{x | R^k_x = R^i_w\}$. ($A_w$ may be empty.)

Fact 5.4.3b (k-node accessibility)**.** If R^i_w is transitive, and reflexive and $N, M \in \mathcal{N}$, then there is an $n \in N$ and $m \in M$ such that $nR^k m$ iff $\forall n \in N, m \in M : nR^k m$. (Why? $\forall m' \in M$: (reflexivity) $m'R^k m'$, so $mR^k m'$, so (transitivity) $nR^k m'$; so $\forall n' \in N : n'R^k m'$.) Thus within R^i_w we can treat R^k as a relation between k-nodes: for $N, M \in \mathcal{N} : NR^k M$ iff $\exists n \in N, m \in M : nR^k m$ iff $\forall n \in N, m \in M : nR^k m$. Similarly for the neighborhood R^k_N of node N. Note: by reflexivity, transitivity, and updating: $NR^k M$ iff $R^k_M \subseteq R^k_N$; and $R^k_M \subset R^k_N$ iff $NR^k M$ and $N \neq M$.

Definition 5.4.3c (maximal k-nodes)**.** Given a transitive, reflexive, and updating neighborhood R^i_w, the *maximal* k-nodes of R^i_w are those who see strictly less than w under R^k but are not seen by any other k-nodes that do so: $\{M \in \mathcal{N} | R^k_M \subset R^i_w$ and $\neg\exists K \in \mathcal{N} : R^k_M \subset R^k_K \subset R^i_w\}$.

Fact 5.4.3d. *If R^i_w is transitive, reflexive, updating, and nested and $M_1, ..., M_n$ are its maximal k-nodes, then it is partitioned by $\{A_w, R^k_{M_1}, ..., R^k_{M_n}\}$.*

Proof. **Exhaustivity:** Take arbitrary $x \in R^i_w$. By definition, $x \in N_x$. If $R^k_{N_x} \not\subset R^i_w$, then by transitivity and updating $R^k_{N_x} = R^i_w$, so $x \in A_w$, hence included in a set in $\{A_w, R^k_{M_1}, ..., R^k_{M_n}\}$. So suppose $R^k_{N_x} \subset R^i_w$; we show that $x \in R^k_{M_j}$ for some maximal M_j. By reflexivity $x \in R^k_x$, so $N_x R^k N_x$. Therefore there must be an node M_j that's maximal and $M_j R^k N_x$. For suppose not: there is no $K \in \mathcal{N}$ such that $R^k_K \subset R^i_w$, $K R^k N_x$, and (by definition of maximal) $\neg\exists K' \in \mathcal{N} : R^k_K \subset R^k_{K'} \subset R^i_w$, i.e.

$$\forall K \in \mathcal{N} : \text{ if } R^k_K \subset R^i_w \text{ and } K R^k N_x \text{ then } \exists K' \in \mathcal{N} : R^k_K \subset R^k_{K'} \subset R^i_w. \qquad (\alpha)$$

But this blows up the size of R^i_w. Since R^i_w is finite, suppose $|R^i_w| = m$. Setting $K = N_x$, we have $R^k_{N_x} \subset R^i_w$ and $N_x R^k N_x$; therefore by (α) there is a K' with $R^k_{N_x} \subset R^k_{K'} \subset R^i_w$. Since $R^k_{N_x} \subset R^k_{K'}$, $K' R^k N_x$. But then setting $K = K'$ we have $R^k_{K'} \subset R^i_w$ and $K' R^k N_x$, so by (α) again we get a K'' such that $R^k_{K'} \subset R^k_{K''} \subset R^i_w$. By iterating this, we prove that $|R^i_w| > m$. Contradiction. Thus there must be a maximal node M_j that accesses N_x, and hence accesses x. Thus $x \in R^k_{M_j}$, as desired.

Exclusivity: If there is an $x \in A_w \cap R^k_{M_j}$, then $M_j R^k A_w$ so by transitivity $R^k_{M_j} \not\subset R^i_w$. Contradiction. So A_w is disjoint from all the $R^k_{M_l}$. Next, take any $M_l \neq M_j$, with $m_l \in M_l$ and $m_j \in M_j$. If $m_l R^k m_j$ or $m_j R^k m_l$, then either they access each other (so by transitivity $M_l = M_j$—contradiction) or only one accesses the other—WLOG, say $m_l R^k m_j$. Since $m_l R^k m_l$ but $m_j \not R^k m_l$, by transitivity $R^k_{m_j} \subset R^k_{m_l} \subset R^i_w$, contradicting the assumption that M_j is maximal. Thus m_l and m_j do not access each other, so by nestedness $R^k_{m_l} \cap R^k_{m_j} = \emptyset$, i.e. $R^k_{M_l}$ and R_{M_j} are disjoint. □

We are now in position to prove the main lemma of Theorem 5.4

Lemma 5.4.3. *If R^i_w is transitive, reflexive, updating, and nested, then at w Naive Trust $\left(P^i(p|P^k(p) \geq t) \geq t\right)$ holds.*

Proof. We will show that Naive Trust holds at w by induction on the size of R^i_w. Note: if $P^i_x(p|q)$ is undefined, $[Pik(p|q) = t]$ holds trivially at x; so to show that Naive Trust holds at x it suffices to show that $P^i_x(p|P^k(p) \geq t) \geq t$ for any p, t on which it's defined.

Base case: $|R^i_w| = 1$, say $R^i_w = \{x\}$. For arbitrary p, t: $P^i_w(p|P^k(p) \geq t)$ is defined iff $P^k_x(p) \geq t$. If so, $P^i_w(P^k(p) \geq t) = 1$, thus $P^i_w(p|P^k(p) \geq t) = P^i_w(p)$. By reflexivity, transitivity, and updating of R^i_w, $R^k_x = \{x\}$; so $P^i_w(p) = \mu(p|R^i_w) = \mu(p|R^k_x) = P^k_x(p) \geq t$, and we have desired result.

Induction step: Suppose $|R^i_w| = l$ and for transitive, reflexive, updating, and nested R^k_x with $|R^k_x| < l$, Naive Trust holds at x. (Since i is a variable, if the hypothesis of Lemma 5.4.3 holds for $|R^i_x| < l$, it holds for both $|R^1_w| < l$ and $|R^2_w| < l$; so we are

allowed to assume it holds for $|R_x^{\boldsymbol{k}}| < l$, not merely $|R_x^i| < l$.) We show that Naive Trust holds at w. By Fact 5.4.3d, since R_w^i is transitive, reflexive, updating, and nested, it can be partitioned into $A_w, R_{M_1}^k, ..., R_{M_n}^k$ for its maximal nodes $M_1, ..., M_n$. Taking arbitrary p, t such that $P_w^i(p|p \geq t)$ is well-defined, either **(i)** $P_w^i(p) < t$ or **(ii)** $P_w^i(p) \geq t$.

Suppose (i): $P_w^i(p) < t$**.** Since $\forall x \in A_w : R_x^k = R_w^i$, it follows that $P^k(p)_x = P_w^i(p) < t$. Hence $A_w \cap [P^k(p) \geq t] = \emptyset$, so $R_{M_1}^k, ..., R_{M_k}^k$ partitions $R_w^i \cap [P^k(p) \geq t]$—the set assigned positive mass by $P_w^i(\cdot|P^k(p) \geq t)$. By the law of total probability,

$$P_w^i(p|P^k(p) \geq t) = \sum_j P_w^i(R_{M_j}^k|P^k(p) \geq t) \cdot P_w^i(p|R_{M_j}^k \cap [P^k(p) \geq t]) \qquad (\alpha)$$

So $P_w^i(p|P^k(p) \geq t)$ is a weighted average of the $P_w^i(p|R_{M_j}^k \cap [P^k(p) \geq t])$. And note that since $R_{M_j}^k \subseteq R_w^i$:

$$\begin{aligned} P_w^i(p|R_{M_j}^k \cap [P^k(p) \geq t]) &= \frac{\mu\big(p \cap R_{M_j}^k \cap [P^k(p) \geq t] \cap R_w^i\big)}{\mu\big(R_{M_j}^k \cap [P^k(p) \geq t] \cap R_w^i\big)} \\ &= \frac{\mu\big(p \cap R_{M_j}^k \cap [P^k(p) \geq t]\big)}{\mu\big(R_{M_j}^k \cap [P^k(p) \geq t]\big)} \\ &= P_{m_j}^k(p|P^k(p) \geq t) \end{aligned}$$

for $m_j \in M_j$. Moreover, since $R_{m_j}^k \subseteq R_w^i$, by Lemma 5.4.2, $R_{m_j}^k$ is transitive, reflexive, updating, and nested. And since $R_{m_j}^k \subset R_w^i$, $|R_{m_j}^k| < |R_w^i| = l$, so by the inductive hypothesis Naive Trust holds at each m_j. In particular, $P_{m_j}^k(p|P^k(p) \geq t) \geq t$ when it's defined $\big($when $P_w^i\big(R_{M_j}^k \cap [P^k(p) \geq t]\big) > 0\big)$. Plugging this into (α):

$$P_w^i(p|P^k(p) \geq t) \;\geq\; t \sum_j P_w^i(R_{M_j}^k|P^k(p) \geq t) = t \cdot 1$$

That is, $P_w^i(p|P^k(p) \geq t) \geq t$, as desired.

Suppose (ii): $P_w^k(p) \geq t$. Then $A_w \cap [P^k(p) < t] = \emptyset$, so by parallel reasoning:

$$P_w^i(p|P^k(p) < t) = \sum_j P_w^i(R_{M_j}^k|P^k(p) < t) \cdot P_w^i\big(p|R_{M_j}^k \cap [P^k(p) < t]\big) \qquad (\beta)$$

And similarly $P_w^i(p|R_{M_j}^k \cap [P^k(p) < t]) = P_{m_j}^k(p|P^k(p) < t) < t$, since the m_j satisfy Naive Trust. Applied to (β) we get $P_w^i(p|P^k(p) < t) < t$. But since $[P^k(p) < t]$ and $[P^k(p) \geq t]$ partition R_w^i and (by hypothesis) $P_w^i(p) \geq t$, we have:

$$t \;\leq\; P_w^i(p) \;=\; P_w^i(P^k(p) < t) \cdot P_w^i(p|P^k(p) < t) \;+\; P_w^i(P^k(p) \geq t) \cdot P_w^i(p|P^k(p) \geq t)$$

So $P_w^i(p|P^k(p) < t)$ and $P_w^i(p|P^k(p) \geq t)$ must average to at least t. Since we know the former is less than t, the latter must be greater: $P_w^i(p|P^k(p) \geq t) \geq t$, as desired.

Since p, t were arbitrary, Naive Trust holds at w—completing the induction. □

Finally, we are in a position to prove our theorem:

Theorem 5.4 (Characterization)**.** *A probabilistic frame $\langle W, R^1, R^2, \mu\rangle$ validates Trust iff $\langle W, R^1, R^2\rangle$ is transitive, surely-reflexive, surely-updating, and surely-nested.*

Proof. ($\Rightarrow$). Supposing $\langle W, R^1, R^2, \mu\rangle$ validates Trust, Lemma 5.4.1 implies that it is transitive, surely-reflexive, surely-updating, and surely-nested.

($\Leftarrow$). Suppose $F = \langle W, R^1, R^2, \mu\rangle$ is transitive, surely-reflexive, surely-updating, and surely-nested. Taking an arbitrary world w, this means R^i_w is transitive, reflexive, updating, and nested. Now take arbitrary q, p, t such that $P^i_w\big(p|q \cap [P^k(p|q) \geq t]\big)$ $\big(= P^i_{w|q}(p|[P^k_q(p) \geq t])\big)$ is defined. Is this value at least t? Consider updating R^i on q to get a new relation R^{i+} such that $xR^{i+}y$ iff xR^iy and $y \in q$; equivalently $R^{i+}_x = q \cap R^i_x$. We can use this to define new probability functions at worlds and propositions about them: $P^{i+}_x =_{df} \mu(\cdot|R^{i+}_x)$, $[P^{i+}(p) = t] =_{df} \{x|P^{i+}_x(p) = t\}$, etc. Likewise define R^{k+} such that $R^{k+}_x = q \cap R^k_x$, with $[P^{k+}(p) = t]$ (etc.) defined in parallel. First note that if $P^k_x(q) > 0$, then

$$P^k_x(p|q) = \mu(p|q \cap R^k_x) = \mu(p|R^{k+}_x) = P^{k+}_x(p) \qquad (\alpha)$$

Since R^i_w is reflexive, every $x \in q \cap R^i_w$ has $P^k_x(q) > 0$; so (α) implies $q \cap R^i_w \cap [P^k(p|q) \geq t] = q \cap R^i_w \cap [P^{k+}(p) \geq t]$. Therefore

$$\begin{aligned} P^i_w\big(p|q \cap [P^k(p|q) \geq t]\big) &= P^i_w\big(p|q \cap R^i_w \cap [P^k(p|q) \geq t]\big) \\ &= P^i_w\big(p|q \cap R^i_w \cap [P^{k+}(p) \geq t]\big) \\ &= P^{i+}_w\big(p|R^i_w \cap [P^{k+}(p) \geq t]\big) \\ &= P^{i+}_w\big(p|P^{k+}(p) \geq t]\big). \end{aligned}$$

Finally, note that by Lemma 5.4.2, $R^{i+}_w = q \cap R^i_w$ is transitive, reflexive, updating, and nested. Therefore by Lemma 5.4.3, Naive Trust holds for P^{i+}_w with respect to P^{k+}: $P^{i+}_w\big(p|P^{k+}(p) \geq t]\big) \geq t$. It follows by our above equality that $P^i_w\big(p|q \cap [P^k(p|q) \geq t]\big) \geq t$, as desired: $P^i_{w|q}(p|P^k_q(p) \geq t) \geq t$. Since w, q, p, t were arbitrary, $\langle W, R^1, R^2, \mu\rangle$ validates (full) Trust. □

Next we move on to characterizing Value.

Remark (Expectations)**.** As we have seen, the expectation of a random variable (function from worlds to numbers) X is defined by $E^i_w[X] =_{df} \sum_s P^i_w\big(X = s\big) \cdot s$. But a more convenient form to work with is given by the total expectation theorem: given any $A_1, ..., A_n$ that partitions R^i_w, $E^i_w[X] = \sum_{A_i} P^i_w(A_i) \cdot E^i_w[X|A_i]$ where $E^i_w[X|A_i]$ is the expectation of X calculated using $P^i_w(\cdot|A_i)$. As a limiting case, we can use the maximally fine-grained partition to get $E^i_w[X] = \sum_{w' \in W} P^i_w(w') \cdot X(w')$. We use this and similar facts about expectations freely in what follows.

Proposition* 6.1. *Value is inconsistent with Improbable Knowing, Misguided Evidence, and Self-Effacing Evidence.*

Proof. **Improbable Knowing:** Suppose $S^i(p \wedge [P^i(S^i p) < \frac{1}{2}])$ at w. We show that Value fails. Let $O = \{n, b\}$ and let

$$U(n, v) = 0 \text{ for all } v \in W \qquad U(b, v) = \begin{cases} 1 & \text{if } v \in p \\ -n & \text{if } v \notin p \end{cases} \quad \text{for large } n > 0.$$

Since $S^i p$ is true at w, $P^i_w(p) = 1$, so $P^i_w(U(b) = 1) = 1$, hence $E^i_w[U(b)] = 1$. But since $P^i_w(S^i p) < \frac{1}{2}$, $P^i_w(P^i(p) < 1) > \frac{1}{2}$. Take such a x seen by w with $P^i_x(p) < 1$. For large enough n, $E^i_x[U(b)] < 0 = E^i_x[U(n)]$, so $D^i_x = n$. Hence $P^i_w(U(D^i) < 1) > 0$ while $P^i_w(U(D^i) \leq 1) = 1$ (since $[U(D^i) \leq 1]$ everywhere), which implies that $E^i_w[U(D^i)] < 1 = E^i_w[U(b)]$. Value fails at w.

Misguided Evidence: Suppose $P^i_w(p \wedge [P^i(p) < t]) \geq t$, so $P^i_w(p \wedge [P^i(p) \leq t - d]) \geq t$ for some $d > 0$. Equivalently, $P^i_w\big(p \wedge [P^i(\neg p) \geq 1 - t + d]\big) \geq t$. Abbreviate $X = p \wedge [P^i(\neg p) \geq 1 - t + d]$. We'll define a decision problem that's a bet on $\neg p$ which isn't worth the risk but which your evidence recommends taking throughout X—where it won't pay out. Let $O = \{n, b\}$. The *n*ope option has 0 utility everywhere, while the *b*et option is a bet on $\neg p$:

$$U(n, v) = 0 \text{ for all } v \in W \qquad U(b, v) = \begin{cases} t & \text{if } v \in \neg p \\ t - 1 - \epsilon & \text{if } v \in p \end{cases} \quad \text{for small } \epsilon > 0.$$

Since $E^i_w[U(n)] = 0$, it'll suffice to show that $E^i_w[U(D^i)] < 0$. At each world $x \in X$, $P^i_x(\neg p) \geq 1 - t + d$, meaning the expected utility of b is $E^i_x[U(b)] = (1 - t + d)(t) + (t - d)(t - 1 - \epsilon) = d + d\epsilon - t\epsilon$. Once $\epsilon < \frac{d}{t}$, this value goes positive: $E^i_x[U(b)] > 0 = E^i_x[U(n)]$. Thus $D^i_x = b$ for every $x \in X$. Recalling that $X \subseteq p$ so $X \cap \neg p = \emptyset$, the bet does not pay out there: $\forall x \in X : U(D^i, x) = U(b, x) = t - 1 - \epsilon$, hence $E^i_w[U(D^i)|X] = t - 1 - \epsilon$. This allows us to derive (2) from (1) below; (3) follows by noting that t is the largest utility obtainable at any world; and (4) follows since $P^i_w(X) \geq t$:

$$\begin{aligned} E^i_w[U(D^i)] &= P^i_w(X) \cdot E^i_w[U(D^i)|X] \;+\; P^i_w(\neg X) \cdot E^i_w[U(D^i)|\neg X] && (1) \\ &= P^i_w(X)(t - 1 - \epsilon) \;+\; P^i_w(\neg X) \cdot E^i_w[U(D^i)|\neg X] && (2) \\ &= P^i_w(X)(t - 1 - \epsilon) \;+\; P^i_w(\neg X)t && (3) \\ &\leq t(t - 1 - \epsilon) \;+\; (1 - t)t \;=\; -t\epsilon \;<\; 0 && (4) \end{aligned}$$

Thus $E^i_w[U(D^i)] < 0 = E^i_w[U(n)]$: Value fails.

Self-Effacing Evidence: Suppose $S^i(p \leftrightarrow [P^i(p) < \frac{1}{2}])$ and $S^i(\neg p \leftrightarrow [P^i(p) > \frac{1}{2}])$ at

w. Let $O = \{n, b_1, b_2\}$ and

$$U(n, v) = 0 \text{ for all } v \in W \qquad U(b_1, v) = \begin{cases} 1 & \text{if } v \in p \\ -1 & \text{if } v \notin p \end{cases} \qquad U(b_2, v) = \begin{cases} -1 & \text{if } v \in p \\ 1 & \text{if } v \notin p \end{cases}$$

Clearly $E^i_w[U(n)] = 0$ since $P^i_w(U(n) = 0) = 1$. So it suffices to show $E^i_w[U(D^i)] < 0$. Take an arbitrary $x \in R^i_w$. Suppose $x \in p$, then $P^i_x(p) < \frac{1}{2}$ meaning that $E^i_x[U(b_2)] > 0 = E^i_x[U(n)] > E^i_x[U(b_1)]$, so $D^i_x = b_2$. Since $x \in p$, b_2 doesn't pay out, so $U(D^i, x) = -1$. Next suppose $x \notin p$, so $P^i_x(p) > \frac{1}{2}$. By parallel reasoning, $D^i_x = b_1$ and so $U(D^i, x) = -1$. Since this applies to any x seen by w, $P^i_w(U(D^i) = -1) = 1$. Thus $E^i_w[U(D^i)] = -1 < 0 = E^i_w[U(n)]$. Value fails at w. Even more: D^i is strictly dominated by n: $S^i(U(D^i) < U(n))$. □

Theorem 6.2: The Value of Evidence Theorem

Recall that Value is validated by $\langle W, R^1, R^2, \mu\rangle$ iff for every decision problem $\langle O, U\rangle$ it satisfies the following inequality for all w, i, k, D, o:

$$\textbf{Value:} \quad E^i_w[U(D^k)] \geq E^i_w[U(o)] \qquad (k \geq i)$$

Here $o \in O$ is an option, while D^k is a function from worlds w to options that maximize expected utility with respect to P^k_w—subject to the constraint that if $P^i_x = P^k_y$, then $D^i_x = D^k_y$. Here is our flagship theorem:

Theorem 6.2 (Value of Evidence Theorem)**.** *The following are equivalent:*

(1) *The probabilistic frame $\langle W, R^1, R^2, \mu\rangle$ validates Trust.*

(2) *$\langle W, R^1, R^2\rangle$ is transitive, surely-reflexive, surely-updating, and surely-nested.*

(3) *The probabilistic frame $\langle W, R^1, R^2, \mu\rangle$ validates Value.*

Theorem 5.4 has already established that (1) and (2) are equivalent, so we must show that (2) and (3) are. We will break it into two stages. As mentioned in §7.1, the result that (2) implies (3) is due in its essentials to Geanakoplos (1989). Here we are in a slightly different framework, which requires slightly different proof methods—I will make use of the tools we developed for Theorem 5.4.

Lemma 6.2.1 (Geanakoplos)**.** *If $\langle W, R^1, R^2\rangle$ is transitive, surely-reflexive, surely-updating, and surely-nested, then $\langle W, R^1, R^2, \mu\rangle$ validates Value.*

Proof. As with Theorem 5.4, it'll suffice to show that if R^i_w is transitive, reflexive, updating, and nested, then Value holds at w for an arbitrary decision problem. We do this by induction on the size of R^i_w.

Base case: $|R^i_w| = 1$, say $R^i_w = \{x\}$. It'll suffice to show that $t = \max_{o \in O} \left(E^i_w[U(o)]\right) \leq E^i_w[U(D^k)]$. By definition, $t = E^i_w[U(D^i_w)]$. But by reflexivity, updating, and transitivity: $R^k_x = \{x\}$, so $P^i_w = P^k_x$, so $D^k_x = D^i_w$. Since this is the only possibility P^i_w assigns positive mass to, $P^i_w\left([U(D^k) = U(D^i_w)]\right) = 1$, which implies our desired result.

Induction step: Suppose $|R^i_w| = n$ and that for each transitive, reflexive, updating, nested R^k_x such that $|R^k_x| < n$, Value holds at x. We'll show it holds at w. Recall from Definition 5.4.3c and Fact 5.4.3d that since R^i_w is transitive, reflexive, updating, and nested, we can take it's maximal k-nodes $M_1, ..., M_l$ and partition it by their neighborhoods plus A_w (which is possibly empty, since perhaps nothing in R^i_w sees all of R^i_w under R^k): $\{A_w, R^k_{M_1}, ..., R^k_{M_l}\}$. Thus taking an arbitrary option o, by the total expectation theorem we have:

$$\begin{aligned} E^i_w[U(o)] &= P^i_w(A_w)E^i_w[U(o)|A_w] + \sum_j P^i_w(R^k_{M_j})E^i_w[U(o)|R^k_{M_j}] \\ &\leq E^i_w[U(D^i_w)] && \text{[By definition of } D\text{]} \\ &= P^i_w(A_w)E^i_w[U(D^i_w)|A_w] + \sum_j P^i_w(R^k_{M_j})E^i_w[U(D^i_w)|R^k_{M_j}] && (\alpha) \end{aligned}$$

So it'll suffice to show that (α) is no greater than $E^i_w[U(D^k)]$. Break this into:

$$E^i_w[U(D^k)] = P^i_w(A_w)E^i_w[U(D^k)|A_w] + \sum_j P^i_w(R^k_{M_j})E^i_w[U(D^k)|R^k_{M_j}] \qquad (\beta)$$

Note that every $x \in A_w$ has $R^k_x = R^i_w$ and therefore $P^k_x = P^i_w$; thus $D^k_x = D^i_w$. Plugging this into the left summand of (β) shows it to be equal to the left summand of (α):

$$P^i_w(A_w)E^i_w[U(D^i_w)|A_w] = P^i_w(A_w)E^i_w[U(D^k)|A_w] \qquad (\gamma)$$

Now we turn to the right summands. Since the M_j are maximal k-nodes, we know each $R^k_{M_j} \subset R^i_w$, thus $|R^k_{M_j}| < |R^i_w| = n$. By Lemma 5.4.2, they are also transitive, reflexive, updating, and nested; so by the inductive hypothesis Value holds at each $m_j \in M_j$; hence for any option $o' \in O$, $E^k_{m_j}[U(o')] \leq E^k_{m_j}[U(D^k)]$. In particular, we can set $o' = D^i_w$ to obtain:

$$E^k_{m_j}[U(D^i_w)] \leq E^k_{m_j}[U(D^k)] \qquad (\delta)$$

Now in general for a random variable X we have

$$\begin{aligned} E^k_{m_j}[X] &= \sum_s P^k_{m_j}(X = s)s \\ &= \sum_s \mu(X = s|R^k_{M_j})s \\ &= \sum_s \mu(X = s|R^k_{M_j} \cap R^i_w)s && \text{[Since } R^k_{M_j} \subseteq R^i_w\text{]} \\ &= \sum_s P^i_w(X = s|R^k_{M_j})s \\ &= E^i_w[X|R^k_{M_j}] \end{aligned}$$

Letting $X = U(D^i_w)$ and then $X = U(D^k)$, respectively, and combining with (δ): for each M_j we obtain $E^i_w[U(D^i_w)|R^k_{M_i}] = E^k_{m_j}[U(D^i_w)] \leq E^k_{m_j}[U(D^k)] = E^i_w[U(D^k)|R^k_{M_i}]$. Plugging this into the right summands of (α) and (β) yields

$$\begin{aligned} \sum_j P^i_w(R^k_{M_j})E^i_w[U(D^i_w)|R^k_{M_j}] &= \sum_j P^i_w(R^k_{M_j})E^k_{m_j}[U(D^i_w)] \\ \leq \sum_j P^i_w(R^k_{M_j})E^k_{m_j}[U(D^k)] &= \sum_j P^i_w(R^k_{M_j})E^i_w[U(D^k)|R^k_{M_j}] && (\epsilon) \end{aligned}$$

Finally, combining $(\alpha), (\beta), (\gamma)$, and (ϵ) yields the desired result: $E^i_w[U(D^k)] \geq E^i_w[U(o)]$. Since o and D were arbitrary, this completes the induction and establishes the result. □

The final step is to show the converse. Though the details are messy, the basic idea is that whenever a frame is not nested, any prior over it will have an "imbalance" in it—a proposition on which it can be expected to (slightly) mislead. By carefully choosing options and utilities to draw out this imbalance, we can find a decision problem on which Value fails.

Lemma 6.2.2. *If $\langle W, R^1, R^2\rangle$ is not transitive, surely-reflexive, surely-updating, and surely-nested, then $\langle W, R^1, R^2, \mu\rangle$ does not validate Value.*

Proof. **Transitivity:** Suppose there is an $x \in R^i_w$ such that there is a $z \in W$ with zR^kx and xR^ky but $z \not R^k y$. Let $O = \{n, b\}$ with

$$U(n, v) = 0 \text{ for all } v \in W \qquad U(b, v) = \begin{cases} 1 & \text{if } v = y \\ -\epsilon & \text{if } v \neq y \end{cases} \text{ for small } \epsilon > 0.$$

Since $P^k_x(y) = P^k_x(U(b) = 1) > 0$, as $\epsilon \to 0$ we get $E^k_x[U(b)] > 0 = E^k_x[U(n)]$. Once this happens, since $P^k_z(y) = P^k_z(U(D^k) > 0) = 0$ (since D^k is 0 or $-\epsilon$ everywhere else), then since $P^k_z(x) \leq P^k_z(U(D^k) < 0) > 0$, we get $E^k_z[U(D^k)] < 0 = E^k_z[U(n)]$. Value fails at z.

Surely-Reflexivity: Suppose $\langle W, R^1, R^2\rangle$ is not surely-reflexive, so there is an $x \in R^i_w$ with $xR\!\!\!/^k x$. We find a decision problem where Value fails. Let $O = \{n, b\}$ ($n$ope and $b$et) with

$$U(n,v) = 0 \text{ for all } v \in W \qquad U(b,v) = \begin{cases} \epsilon & \text{if } v \neq x \\ -1 & \text{if } v = x \end{cases} \quad \text{for small } \epsilon > 0.$$

Since $xR\!\!\!/^k x$, $P^k_x(x) = 0$, so $E^k_x[U(b)] = \epsilon > 0 = E^k_x[U(n)]$, hence $D^k_x = b$. But since $wR^i x$, $P^i_w(x) > 0$; since $[U(D^k) = -1]$ at x, $P^k_w(U(D^k) = -1) = a > 0$. Thus as $\epsilon \to 0$ (in particular, $\epsilon < a$) we obtain $E^i_w[U(D^k)] < 0 = E^i_w[U(n)]$. Value fails at w.

Surely-Updating: We know R^i_w must be surely-reflexive. Suppose there is an $x \in R^i_w$ with $R^2_x \not\subseteq R^1_x$; say xR^2y but $xR\!\!\!/^1 y$. By surely-reflexivity, xR^1x. Let $O = \{n, b\}$ with:

$$U(n,v) = 0 \text{ for all } v \in W \qquad U(b,v) = \begin{cases} 1 & \text{if } v = y \\ -\epsilon & \text{if } v \neq y \end{cases} \quad \text{for small } \epsilon > 0.$$

Since $P^2_x(y) = P^2_x(U(b) = 1) > 0$, as $\epsilon \to 0$ we get $E^2_x[U(b)] > 0 = E^2_x[U(n)]$, hence $D^2_x = b$. Now $P^1_x(y) = P^1_x(U(D^2) > 0) = 0$ (since D^k is 0 or $-\epsilon$ everyone else); so since $0 < P^1_x(x) \leq P^1_x(U(D^2) < 0)$, we have $E^1_x[U(D^2)] < 0 = E^1_x[U(n)]$. Value fails at x.

Nestedness: We know R^i_w must be transitive, surely-reflexive, and updating. Suppose it's not nested: for $x, y \in R^i_w$: $xR\!\!\!/^k y$ and $yR\!\!\!/^k x$ but $R^k_x \cap R^k_y \neq \emptyset$. Recalling that $N_v = \{v' \in R^i_w | R^k_{v'} = R^k_v\}$, define $C = (R^k_x \cup R^k_y) - (N_x \cup N_y)$. This is the proposition we will "bet" on. Basically, since R^k_x and R^k_y overlap, C "looms larger" to them than it should according to w—who can see the whole setup.

Before defining our decision problem, we record some facts about this frame. (1) $R^k_x \cup R^k_y \subseteq R^i_w$ (transitivity, updating). (2) $R^k_x \cap R^k_y \subseteq C$, for otherwise they'd overlap in $N_x \cup N_y$ and so see each other. (3) For $v \in C : P^k_v(C) = 1$ (by transitivity, $R^k_v \subseteq R^k_x \cup R^k_y$; and if v saw one of x or y, either $v \in N_x, N_y, xR^k y$, or $yR^k x$—all contradictions). (4) Since $x \in R^k_x$ and $y \in R^k_y$ (reflexivity), $0 < P^k_x(C), P^k_y(C) < 1$. Finally, supposing v is x or y, v' is the other, and $P^k_v(C) = t$, then $(\boldsymbol{\alpha})$: $P^i_w(C|R^k_v - R^k_{v'}) < t$. For since $R^k_v \subseteq R^i_w$, $P^i_w(C|R^k_v) = \mu(C|R^k_v) = P^k_v(C) = t$, and by total probability:

$$= \mu(R^k_{v'}|R^k_v)\mu(C|R^k_x \cap R^k_y) + \mu(R^k_v - R^k_{v'}|R^k_v)\mu(C|R^k_v - R^k_{v'})$$

So $\mu(C|R^k_x \cap R^k_y)$ and $\mu(C|R^k_v - R^k_{v'})$ average to $t < 1$. By (2) the first equals 1, so the second must be less than t.

We can now set up our decision problem. Without loss of generality, suppose

$P_y^k(C) \leq P_x^k(C)$, so $P_y^k(C) = t \leq P_x^k(C) = t + d$ for $d \geq 0$. Let $O = \{n, b\}$ with

$$U(n,v) = \begin{cases} 0 & \text{if } v \notin N_x \\ \frac{d}{1-t-d} & \text{if } v \in N_x \end{cases} \qquad U(b,v) = \begin{cases} 1-t+\epsilon & \text{if } v \in C \\ -t & \text{if } v \notin C \end{cases} \quad \text{for small } \epsilon > 0.$$

First note that, again, $(\boldsymbol{\beta})$ **:** $\forall v \in R_x^k \cup R_y^k = C \cup N_x \cup N_y : D_v^k = b$. For if $v \in C$, $P_v^k(C) = 1$ and $P_v^k(N_x) = 0$, so $E_v^k[U(n)] = 0 < 1 - t + \epsilon = E_v^k[U(b)]$. If $v \in N_y$, then $P_v^k(N_x) = 0$ while $P_v^k(C) = t$, so $E_v^k[U(b)] = t(1-t+\epsilon)+(1-t)(-t) = t\epsilon > 0 = E_v^k[U(n)]$. Finally, if $v \in N_x$ then

$$\begin{aligned} E_v^k[U(n)] &= P_x^k(C)0 + P_x^k(\neg C)\frac{d}{1-t-d} = (1-t-d)\frac{d}{1-t-d} = d \\ &< E_v^k[U(b)] = P_x^k(C)(1-t+\epsilon) + P_x^k(\neg C)(-t) \\ &= (t+d)(1-t+\epsilon) + (1-t-d)(-t) = d + t\epsilon + d\epsilon. \end{aligned}$$

We next show that $E_w^i[U(b)|R_x^k \cup R_y^k] < E_w^i[U(n)|R_x^k \cup R_y^k]$.

Proof. By total expectation we have $E_w^i\big[U(n)\big|R_x^k \cup R_y^k\big]$

$$= P_w^i(R_x^k|R_x^k \cup R_y^k)E_w^i[U(n)|R_x^k] + P_x^k(R_y^k - R_x^k\big|R_x^k \cup R_y^k)E_w^i[U(n)|R_y^k - R_x^k].$$

Since $\forall v \in R_y^k - R_x^k : U(n,v) = 0$, the second summand is 0. Since $R_x^k \subseteq R_w^i$, $E_w^i[U(n)|R_x^k] = E_x[U(n)] = d$. Therefore $(\boldsymbol{\gamma})$ **:** $E_w^i[U(n)|R_x^k \cup R_y^k] = P_w^i(R_x^k|R_x^k \cup R_y^k)d$.

On the other hand, $E_w^i[U(b)|R_x^k \cup R_y^k]$

$$\begin{aligned} &= P_w^i(R_x^k|R_x^k \cup R_y^k)E_w^i[U(b)|R_x^k] + P_w^i(R_y^k - R_x^k|R_x^k \cup R_y^k)E_w^i[U(b)|R_y^k - R_x^k] \\ &= P_w^i(R_x^k|R_x^k \cup R_y^k)(d + t\epsilon + d\epsilon) + P_w^i(R_y^k - R_x^k|R_x^k \cup R_y^k)E_w^i[U(b)|R_y^k - R_x^k] \end{aligned}$$

Taking out a $P_w^i(R_x^k|R_x^k \cup R_y^k)d$ to subtract from both this and (γ), and factoring, it suffices to show that

$$0 > \epsilon P_w^i(R_x^k|R_x^k \cup R_y^k)(t+d) + P_w^i(R_y^k - R_x^k|R_x^k \cup R_y^k)E_w^i[U(b)|R_y^k - R_x^k]$$

Since the left summand approaches 0 as ϵ does, it in turn suffices to show that $E_w^i[U(b)|R_y^k - R_x^k] < 0$ for small ϵ. Recall that by (α), $P_w^i(C|R_y^k - R_x^k) < t$, so it equals $t - b$ for $b > 0$. Thus

$$E_w^i[U(b)|R_y^k - R_x^k] = (t-b)(1-t+\epsilon) + (1-t+b)(-t) = t\epsilon - b - b\epsilon.$$

As $\epsilon \to 0$ we have this goes negative: $E_w^i[U(b)|R_y^k - R_x^k] < 0$ for small ϵ, as desired. Hence $(\boldsymbol{\delta})$ **:** $E_w^i\big[U(b)\big|R_x^k \cup R_y^k\big] < E_w^i\big[U(n)\big|R_x^k \cup R_y^k\big]$ □

Now by the total expectation, compare:

$$E_w^i[U(D^k)] = P_w^i(R_x^k \cup R_y^k)E_x[U(D^k)|R_x^k \cup R_y^k] + P_w^i(\neg(R_x^k \cup R_y^k))E_w^i[U(D^k)|\neg(R_x^k \cup R_y^k)] \quad (1)$$

$$E_w^i[U(n)] = P_w^i(R_x^k \cup R_y^k)E_x[U(n)|R_x^k \cup R_y^k] + P_w^i(\neg(R_x^k \cup R_y^k))E_w^i[U(n)|\neg(R_x^k \cup R_y^k)] \quad (2)$$

Outside $R^k_x \cup R^k_y$, $U(n, v) = 0$ and $U(D^k, v) \leq 0$, so the second summand of (1) is $\leq$ that of (2). But combining (β) with (δ) yields $E^i_w[U(D^k)|R^k_x \cup R^k_y] = E^i_w[U(b)|R^k_x \cup R^k_y] < E^i_w[U(n)|R^k_x \cup R^k_y]$, so the left summand of (1) is $<$ that of (2), meaning $E^i_w[U(D^k)] < E^i_w[U(n)]$. Value fails at w. □

Our flagship awaits:

Theorem 6.2 (Value of Evidence Theorem)**.** *The following are equivalent:*

(1) *The probabilistic frame* $\langle W, R^1, R^2, \mu\rangle$ *validates Trust.*

(2) $\langle W, R^1, R^2\rangle$ *is transitive, surely-reflexive, surely-updating, and surely-nested.*

(3) *The probabilistic frame* $\langle W, R^1, R^2, \mu\rangle$ *validates Value.*

Proof. By Theorem 5.4, (1) holds iff (2) does. By Lemma 6.2.1, if (2) holds then (3) does. And by Lemma 6.2.2, if (2) does not hold then (3) does not. Combined, we have the result. □