

Evidence: A Guide for the Uncertain

KEVIN DORST

Massachusetts Institute of Technology

Assume that it is your *evidence* that determines what opinions you should have. I argue that since you should take peer disagreement seriously, evidence must have two features. (1) It must sometimes warrant being *modest*: uncertain what your evidence warrants, and (thus) uncertain whether you're rational. (2) But it must always warrant being *guided*: disposed to treat your evidence as a guide. It is surprisingly difficult to vindicate these dual constraints. But diagnosing why this is so leads to a proposal—Trust—that is weak enough to allow modesty but strong enough to yield many guiding features. In fact, I argue that Trust is the *Goldilocks principle*—for it is necessary and sufficient to vindicate the claim that you should always prefer to use free evidence. Upshot: Trust lays the foundations for a theory of disagreement and, more generally, an epistemology that permits self-doubt—a modest epistemology.

1. A Modest Guide

Here is a spoon:



And here is another:



Which spoon is longer? Let *Top* be the claim that the top one is. There are opinions you should have about *Top*—perhaps you should be confident of it. There are opinions you should have about many other propositions as well. I will assume that it is your *evidence* that makes it so. But I am neutral beyond that—all I assume is that your evidence determines what opinions you should have. Thus: You should (would be rational to) have an

opinion iff your evidence warrants having that opinion.¹ Epistemology studies evidence. And this paper studies its structure.

Two structural features, in fact.

One. Your evidence determines what *you* ought to think. Yet you are a fallible critter—and you know it. Even in simple cases, you might misjudge the force of your evidence: although you're pretty confident of *Top* (let's suppose), maybe it should be obvious that the top spoon is longer (you're underconfident)—or maybe you're being tricked by its shape (you're overconfident). You're aware of your fallibility, so you have evidence that requires being unsure whether you've properly conformed to your evidence. That's our first structural feature:

MODEST TRUISM

Your evidence sometimes requires being *modest*: uncertain what your evidence requires.

Let '*P*' be a definite description that picks out the credences—whatever they are—that are warranted by your evidence. *P* encodes the opinions that you should have about any relevant proposition; thus it captures the first-order opinions you should have about *Top*, as well as the *higher-order opinions* you should have about what first-order opinions you should have about *Top*. Our Modest Truism says that such first- and higher-order opinions should sometimes come apart—sometimes you should have credence *t* in *p* but be less than certain that you should: $P(p) = t$, but $P(P(p) = t) < 1$. Maybe you should be 0.7 confident of *Top*: $P(\text{Top}) = 0.7$. But you definitely shouldn't be *certain* that you should be—you should leave open that you should be 0.6 or 0.8 instead: $P(P(\text{Top}) = 0.6) > 0$ and $P(P(\text{Top}) = 0.8) > 0$.²

Two. Your evidence determines what you *ought* to think. Arguably, this means that you must have reason to treat your evidence as a guide: your higher-order opinions about what you *ought* to think about *Top* should constrain your first-order opinions about *Top*. That's our second (not yet fully precise) structural feature:

GUIDING TRUISM

Your evidence always warrants being *guided*: disposed to treat your evidence as a guide.

There are many things you should be disposed to treat as guides—chances (Lewis 1980), gurus (Elga 2007), future selves (van Fraassen 1984). But evidence must be a *modest* guide—a guide that is not sure that it is a guide. That's...

¹ *Assumption Alert*: (1) I will use precise probabilities to model rational opinions (cf. White 2009a; Elga 2010; Joyce 2010); and (2) I will assume that your evidence always warrants a unique opinion (cf. White 2005; Feldman 2007; Schoenfield 2014). These are modeling choices: (1) *however* we represent rational opinions, it will be rational to be unsure what the rational opinion is; and (2) *whatever* features (evidence, priors, etc.) collectively determine a uniquely rational opinion, they will need the structural features I discuss. I understand these assumptions to be *descriptive* idealizations—like point-particles on frictionless planes—not normative ones. They are intended to provide simplified models of how *you* and *I* ought to think, rather than exact models of how our ideal counterparts would.

² Our Modest Truism allows rational agents to know what their credences are. Let '*C*' be a definite description for your actual credences. Then you may in fact be rational ($[C = P]$ is true) and be certain of what your actual credences are ($[C(p) = t] \rightarrow [C(C(p) = t) = 1]$ holds), but be unsure whether your actual credences are rational: $C(C = P) < 1$. I will suppress talk of *C*, since it is the rational credences *P* (and so, primarily, wide-scope norms) we are interested in. Note: '*P*' is descriptive even when embedded in larger constructions; ' $P(P(p) = 0.5) = 0.3$ ' is to be read, 'The rational credence function (whatever it is) assigns 0.3 credence to the claim that the rational credence function (whatever it is) assigns 0.5 credence to *p*.'

The Problem: We must vindicate both our Modest and Guiding Truisms. This is a problem of modulation: our Modest Truism allows you to leave open various possibilities for what the rational opinions might be; our Guiding Truism requires that these various possible rational opinions are correlated with truth. It's a hard problem. Many have argued that it can't be solved—that there's no principled center of gravity that allows modesty but requires guidance.³ The “higher-order evidence” literature contains three responses to this challenge. *Splitters* accept our Modest Truism and so deny our Guiding one—they allow your first- and higher-order opinions to split radically apart.⁴ *Mergers* accept our Guiding Truism and so deny our Modest one—they require you to be certain of what you should think.⁵ *Bridgers* try to find a middle way—they search for principles weak enough to vindicate our Modest Truism but strong enough to vindicate our Guiding one.⁶ Short story: Bridging is in trouble—the current proposals are all either too strong or too weak (Lasonen-Aarnio 2014, 2015; Dorst 2018b).

The Project: Show that Bridging succeeds, after all. There *is* a principled center of gravity that allows modesty while requiring guidance. We *can* construct an epistemology that makes room for modest critters—like you and me—that have a rational dose of self-doubt.

The Plan: After making the case that we should want a Bridging principle (§2), I'll explain why the standard one is too strong (§3) and propose a fix (§4). The rest of the paper argues that this is the correct fix: it has the modest and guiding features we're after (§5), avoids the paradoxes of weaker proposals (§6), coincides with an independent characterization of our Guiding Truism (§7), and promises fruitful applications (§8).

2. A Disagreement

I've offered an intuitive narrative in favor of Bridging. I'll now offer a precise argument: if you should take peer disagreement seriously, our Modest and Guiding Truisms must both be true. Thus Bridging lies at the foundations of the epistemology of disagreement.

Take a paradigm case. Looking at the spoons, you and your colleague Disa know that you share *Top*-related evidence, and so should have the same opinion about *Top*. (If this seems implausible, stipulate that the Epistemology Oracle has announced it.) Suppose that you start out confident of *Top*—and that, in fact, such confidence is warranted by the evidence. Now you discover a disagreement: Disa is *not* confident of *Top*. Claim: this should make you less confident of *Top*. More generally: if you're rational and know that your peer has the same relevant evidence, then your confidence should drop when you learn that she's less confident than you.⁷

³ Smithies (2012, 2015); Lasonen-Aarnio (2014, 2015); Titelbaum (2015); Salow (2017); Horowitz (2018).

⁴ Williamson (2000, 2014); Lasonen-Aarnio (2010, 2014, 2015); Coates (2012); Hazlett (2012); Wedgwood (2012).

⁵ Smithies (2012, 2015); Greco (2014a); Titelbaum (2015); Salow (2017); cf. Tal (2018).

⁶ Feldman (2005); Gibbons (2006); Elga (2007, 2013); White (2009b); Christensen (2010a,b, 2016); Huemer (2011); Horowitz (2014); Pettigrew and Titelbaum (2014); Vavova (2014, 2016); Littlejohn (2015); Schoenfield (2015a,b, 2016); Sliwa and Horowitz (2015); Worsnip (2015).

⁷ This claim presupposes only that—in such highly circumscribed cases—you should not *completely ignore* your peer's lack of confidence. Although Right Reasons theorists may object (Titelbaum 2015), most others will agree (e.g. Christensen 2007, 2010a; Elga 2007; Feldman 2007; Roush 2009; Kelly 2010; Lasonen-Aarnio 2013; Vavova 2014). Moreover, my claim is compatible with “synergistic” views (Easwaran et al. 2016) since what you are learning is that Disa is *less* confident than you—not that she has a particular credence which is slightly less.

First fact: this requires that you be modest. To see why, suppose you were *immodest*: you were confident of *Top*, and were certain that you should be. Since you know that you share *Top*-related evidence, you'd infer that *Disa's* evidence warrants confidence in *Top*. So you'd think, "Disa's not confident of *Top*. But given her evidence, she *should* be. Since her only route to the truth is through her evidence (the evidence includes all facts that determine what she should think, after all), her lack of confidence doesn't tell me anything about *Top*. So I should simply ignore her unwarranted opinion and maintain my confidence in *Top*." Thus if you were immodest, you'd ignore her opinion. Since you *shouldn't* ignore her opinion, you *shouldn't* be immodest. That's our Modest Truism.

More rigorously, let ' P_D ' and ' C_D ' be definite descriptions for the credences Disa should have and in fact has, respectively. Your evidence warrants being sure that you and Disa ought to have the same opinion about *Top*, i.e. $P(P(\textit{Top}) = P_D(\textit{Top})) = 1$. Since the only way Disa can connect her credence with the truth is through her evidence⁸, once you know which opinion her evidence warrants, further learning⁹ her *actual* opinion won't affect the probability of *Top*:

$$P(\textit{Top}|P_D(\textit{Top}) = s) = P(\textit{Top}[P_D(\textit{Top}) = s] \wedge [C_D(\textit{Top}) = t]).$$

Finally, you should leave open that Disa may have a lower credence than you: $P(C_D(\textit{Top}) < 0.7) > 0$. It follows that if you should be immodest ($P(P(\textit{Top}) = 0.7) = 1$), then you should ignore Disa's lack of confidence: $P(\textit{Top}|C_D(\textit{Top}) < 0.7) = P(\textit{Top})$.¹⁰ Since you shouldn't ignore her lack of confidence, you shouldn't be immodest.

So suppose you are modest: you are uncertain what your evidence warrants. Since Disa is a peer, learning that she's not confident of *Top* gives you reason to think that your shared evidence *doesn't* warrant confidence—despite what you initially thought. Second fact: in order for this change to require you to lower your credence in *Top*, you must be guided by your evidence. For if your opinions about your evidence didn't constrain your opinions about *Top*, then the change in higher-order opinion induced by Disa's disagreement needn't lead to a change in first-order opinion.

To illustrate, consider one way our Guiding Truism could fail. Suppose your evidence could warrant the following "akratic" state: being confident that *Top is true, but I shouldn't be confident of it*; $P(\textit{Top} \wedge [P(\textit{Top}) < 0.7]) \geq 0.7$.¹¹ Then even if Disa's disagreement made you *certain* that you shouldn't be confident of *Top*, you should still maintain your confidence in *Top*—after all, you're confident that possibilities where you shouldn't be confident in *Top* are ones where it's true!¹² Generalizing our puzzle:

Misguided Evidence: $\exists p, t : P(p \wedge [P(p) < t]) \geq t$

You should have confidence that: *p* but I shouldn't have confidence that *p*.

⁸ Titelbaum and Kopec (2017) and Levinstein (2017) give cases where your peer has non-evidential routes to the truth, but we can stipulate that our case is one where you know this isn't so.

⁹ Strictly, all the principles I will discuss are about conditional beliefs, not about learning—but for ease of exposition I'll elide this distinction.

¹⁰ $P(\textit{Top}|C_D(\textit{Top}) < 0.7)$ is an average of $P(\textit{Top}|C_D(\textit{Top}) = t)$ for the various $t < 0.7$. Since $P(P(\textit{Top}) = 0.7) = 1$, we know that each such $P(\textit{Top}|C_D(\textit{Top}) = t) = P(\textit{Top}[P(\textit{Top}) = 0.7] \wedge [C_D(\textit{Top}) = t]) = P(\textit{Top}[P_D(\textit{Top}) = 0.7] \wedge [C_D(\textit{Top}) = t]) = P(\textit{Top}|P_D(\textit{Top}) = 0.7) = P(\textit{Top}|P(\textit{Top}) = 0.7) = P(\textit{Top})$.

¹¹ This is a probabilistic version of what many have called "epistemic akrasia"—the paradigm instance being a belief in *p* but I shouldn't believe it (Smithies 2012; Horowitz 2014; Titelbaum 2015). But there are reasons to think that akrasia is not the distinctive feature of such attitudes (Dorst 2018a).

¹² If $P(\textit{Top} \wedge [P(\textit{Top}) < 0.7]) \geq 0.7$, then $P(\textit{Top}|P(\textit{Top}) < .7) = \frac{P(\textit{Top} \wedge [P(\textit{Top}) < 0.7])}{P(P(\textit{Top}) < .7)} \geq 0.7$.

If Misguided Evidence were possible, then in a parallel case (in which you start off *t*-confident and then discover that Disa's credence is lower), you needn't lower your credence in *p*. So if you must *always* take such disagreement seriously, Misguided Evidence must be impossible. More generally, your opinions about whether your evidence supports *p* must always constrain your opinions about *p*. That's our Guiding Truism.

Upshot: To account for the force of peer disagreement, we must build a Bridging theory that vindicates both our Modest and Guiding Truisms. Construction ahead.

3. A Reflection

Looking at the spoons, you should be modest: there are various epistemic states that you should leave open might be the rational one. Here is a metaphorical (and literal) way to conceptualize this scenario (cf. Elga 2013; Hall 1994). Imagine you are a member of an epistemic panel: a group of candidates who share your evidence and have different opinions in response. (Literally: a set of credence functions.) Everyone knows what everyone else's opinions are. (Literally: each credence function is certain of the values of the other functions.) One of the candidates—the most diligent one—is *the expert*. (Literally: one of the credence functions is the one warranted by the evidence.) You *should* have an opinion iff the expert *does* have that opinion. (Literally: you should have an opinion iff your evidence warrants that opinion.) But by our Modest Truism, the expert may be unsure who the expert is—the most diligent person on the panel may not *know* they're the most diligent person on the panel. (Literally: the credence function warranted by the evidence may be uncertain which credence function is warranted by the evidence.) Different candidates have different opinions about who the expert might be: you're sure it's either Disa, Carl, or yourself; Carl is sure it's either Disa or Betty, etc. (Literally: different credence functions assign different probabilities to claims about who the expert is.) Yet—by our Guiding Truism—all the candidates will defer when they learn about the expert's opinions. (Literally: each credence function will have conditional probabilities that defer to facts about the evidence.) The question is how.

Our story begins with the (seemingly) obvious answer. I'll say that the expert has a given **opinion**¹³ about *p* iff their credence falls in a contextually specified range: $P(p) \in [l, h]$. Obvious answer: *defer to expert opinions* (cf. Skyrms 1980; van Fraassen 1984; Gaifman 1988; Christensen 2010b). That is, your opinions should be a reflection of the expert's: conditional on the expert being 0.6 confident of *Top*, be 0.6 confident of it; conditional on the expert being between 0.5 and 0.8 confident of *Top*, be between 0.5 and 0.8 confident of it; and so on. Letting $P(p|q)$ be the rational credence in *p* conditional on *q*:

Reflection: $P(p|P(p) \in [l, h]) \in [l, h]$ ¹⁴

Upon learning that your evidence warrants a given opinion, have that opinion.

Slogan: *defer to expert opinions*.

¹³ *Convention:* technical terms to be used in the statements of principles and theorems are bolded when defined; their definitions are collected in Appendix C.

¹⁴ Look strange? Some might expect to see different probability functions on the inside and outside, e.g. with your actual credences *C* deferring to the rational credences: $C(p|P(p) \in [l, h]) \in [l, h]$. But that principle is false—your *actual* credences can be whatever you like. What's true is that you *ought* to have credences that obey this principle; letting ' $\Box q$ ' mean 'it ought to be that *q*': $\Box(C(p|P(p) \in [l, h]) \in [l, h])$. But your credences ought to be the rational credences: $\Box(C(p|q) = t) \Leftrightarrow [P(p|q) = t]$; hence our principle.

Reflection says to treat the expert as a guide by simply adopting whatever opinions you find out they have. It clearly vindicates our Guiding Truism. What could go wrong?

Modesty could. Reflection is inconsistent with our Modest Truism: it requires you to be certain that the expert is *immodest*. Consider an example. Looking at the spoons, you have two people on your panel—Imani and yourself. Imani is immodest: she's certain she's the expert. You are modest: you're 50-50 on whether you or Imani is the expert. This simple case is inconsistent with Reflection. For conditional on the expert being 0.5 confident that Imani's the expert, how confident should you be that Imani's the expert? Reflection says to adopt the expert's credences: be 0.5. But that's wrong. Imani is certain that she's the expert. You are 0.5 confident that she is. So if the expert (whoever it is) is 0.5 confident that Imani's the expert, then the expert (whoever it is) is *not* Imani—it's you. Conditional on the expert being 0.5 confident that Imani's the expert, you should have credence 0 that she is: $P(\text{Imani} | P(\text{Imani}) = 0.5) = 0$. Reflection fails.

Surprisingly, the example generalizes completely: Reflection requires you to be certain that you ought to be immodest. Letting *Immodest* be the proposition that the rational credence function is certain of what the rational credence function is:

Fact 3.1: *If a probability frame validates Reflection, it validates $[P(\text{Immodest}) = 1]$.*¹⁵

Upshot: although Reflection vindicates our Guiding Truism, it is incompatible with our Modest one. A weaker principle is needed.

4. On Trust

Fact 3.1 shows *that* Reflection is too strong. But—in order to refine Reflection—we need to know *why* it's too strong. There are two parts to the explanation.

First: if you should be uncertain what your evidence warrants, then learning facts about your evidence can give you *new* evidence—and so can *change* what it's rational to think (Here I follow Elga 2013). This is what happens in the Imani case: when you learn that the expert is 0.5 confident that Imani's the expert, you learn that *you* are the expert. When the expert (i.e. you) formed their 0.5 credence, they didn't know that. So after learning something about the expert's opinions, you now have information that the expert didn't have when they formed those opinions. Thus you should react to this information *not* by adopting the opinion they *had*, but rather by adopting the opinion they *would* have *were they to learn what you've learned*. When you know more than the expert, you should react to your information *as you know the expert would*.

But when the expert learns that their opinion about *p* was rational, why would that lead them to *change* their opinion? This is the second part of the explanation: the expert will think that *the rational credence in p is correlated with the truth-value of p*, and therefore learning that their opinions were rational can sometimes be evidence *against p*. This implies that Reflection must fail.

¹⁵ A **probability frame** is a structure for modeling the opinions—including higher-order opinions—that a given agent should have in a given scenario. Details are in Appendix A. A probability frame **validates** a principle iff the principle is true at all worlds for all well-defined instantiations of its free variables. Samet (1997), Williamson (2000, 2014), Elga (2013), and Dorst (2018a) prove results similar to Fact 3.1.

Return to the Imani case. Letting p be the proposition that Imani's the expert, the expert is in fact 0.5 confident of p : $P(p) = 0.5$. Moreover, the expert is certain that the expert's credence in p (whatever it is) is either 0.5 or 1: $P([P(p) = 0.5] \vee [P(p) = 1]) = 1$.¹⁶

The expert will think that the rational credence is correlated with truth—so since there are only two possible values of the rational credence (0.5 and 1), they will think that p is more likely to be true if the rational credence is higher than if it's lower:

$$P(p|P(p) = 0.5) < P(p|P(p) = 1)$$

But this implies a Reflection failure. Since the expert's original opinion $P(p)$ is an average of the conditional opinions $P(p|P(p) = 0.5)$ and $P(p|P(p) = 1)$, it follows that upon learning that the rational credence in p was the lower value, the expert will *drop* their credence below its original value of 0.5:

$$P(p|P(p) = 0.5) < P(p) = 0.5$$

In short: learning that the expert has a given opinion about p can sometimes be evidence that the rational credence is *lower* than you expected it to be, and therefore—since rational credence is correlated with truth—can be evidence *against* p . (Similarly for any upper-bounded opinion $P(p) \in [l, h]$ —in the Imani case learning that the rational credence in p is between 0.5 and 0.8 would *also* be evidence against p : $P(p|P(p) \in [0.5, 0.8]) < P(p) = 0.5$.)

So Reflection fails because learning that the expert has a given opinion about p can sometimes be evidence against p . To refine it, we must find a type of information about the rational credence in p that is *never* evidence against p . What could it be?

Consider the claim that the expert is *at least t-confident* of p : $P(p) \geq t$. When this is so, I'll say that the expert **judges** that p (to the contextually specified degree t). We could say that judgments are special cases of opinions ($P(p) \in [l, h]$ where $h = 1$). Or we could say that opinions are conjunctions of judgments.¹⁷ However we carve it up, judgments are important. Why?

Because the claim that the rational credence in p is at least t can only provide evidence that the rational credence is *higher* than you originally thought, and so—since rational credence is correlated with truth—can *never* be evidence against p .

More precisely, suppose—for reductio—that learning that the expert judges p provides evidence *against* p :

$$P(p|P(p) \geq t) < P(p)$$

From this it follows (by total probability) that learning that the rational credence is *at least t* should lead you to *lower* your credence in p , while learning that the rational credence is *less than t* should lead you to *raise* your credence in p :

¹⁶ For aficionados: *this* is the point at which the argument I'm about to give would fail as an argument against (say) the Principal Principle (Lewis 1980). Since the Principal Principle connects two different probability functions—the rational credence and chance—it can forbid the possibility that the rational credence equals the lowest possible chance. However, principles that connect rational credence *with rational credence* must allow that whenever you know what the (finite) range of possible rational credences is, there will be a possibility where the actual rational credence is the lowest possible rational credence—as in our case where $P(p) = 0.5$ while $P([P(p) = 0.5] \vee [P(p) = 1]) = 1$.

¹⁷ Since $P(p) = 1 - P(\neg p)$, an $[l, h]$ -opinion that p is equivalent to the conjunction of an l -judgment that p and a $(1 - h)$ -judgment that $\neg p$: $P(p) \in [l, h] \Leftrightarrow ([P(p) \geq l] \wedge [P(\neg p) \geq 1 - h])$.

$$P(p|P(p) \geq t) < P(p) < P(p|P(p) < t)$$

In other words, you should think that possibilities where the rational credence in p is *higher* are *less* likely to be ones where p is true—you should think that the rational credence is *not* correlated with truth! But that's wrong. Contraposing: since you should think that rational credence *is* correlated with truth, learning that the rational credence in p is above a given threshold can never provide evidence against p . Precisely:

$$P(p|P(p) \geq t) \geq P(p)$$

This is the crux of our story, for it shows us how to refine Reflection.¹⁸ Suppose you learn that the expert judges that p : $P(p) \geq t$. You should react to this information as you know the expert would—so how would the expert react? You've learned that they were originally at least t -confident of p . By the above reasoning, if *they* were to learn what you learned (namely, that the expert judges that p) this wouldn't provide them with any evidence against p —*their credence wouldn't drop*. Since you know that they were originally at least t -confident of p , you can infer that upon learning what you've learned the expert would react by *still* being at least t -confident of p . You should react to your information as you know the expert would. So *you* should be at least t -confident of p . Precisely:

SIMPLE TRUST: $P(p|P(p) \geq t) \geq t$

Upon learning that your evidence warrants judging that p , judge that p .

Slogan: *take expert judgments on trust.*

Simple Trust is what we get when we restrict Reflection to apply to expert opinions of the form $[t,1]$ —opinions that are never evidence against p .

Summing up: For evidence to be a modest guide, rational credence must be correlated with truth. This means two things. First, it means that learning that the expert's credence in p falls *within a range* can be evidence *against* p . That is why Reflection fails. Second, it means that learning that the expert's credence in p falls *above a threshold* can *never* be evidence against p . That's why Simple Trust holds. The rest of the paper defends this solution.

Or rather: a solution *like* Simple Trust. We need one final piece of bookkeeping—one that applies equally to Reflection. No matter what bit of information q you learn, you should *still* treat your evidence (updated on q) as a guide. Thus our deference principles should apply not only to the unconditional attitudes warranted by the evidence, but also the conditional ones. Let $P_q(p)$ be the credence (whatever it is) that your evidence warrants having in p *conditional on* q . Our deference principles should apply with ' P_q ' substituted for ' P '. Of course, doing so yields our original principles as a special case (let $q = p \vee \neg p$); so since Reflection is already too strong, so too is its generalization.

But for principles that are *not* too strong—like Simple Trust—this generalization is exactly what we need. Thus we arrive at the promised principle:

Trust: $P_q(p|P_q(p) \geq t) \geq t$

Upon learning that your evidence warrants reacting to q by judging that p , react to q by judging that p .

Slogan: *take expert judgments on trust.*

¹⁸ The following line of reasoning is formalized in §5.

I claim that Trust is the key to making evidence a modest guide. But I won't ask you to take it on trust—the rest of this paper makes the case.

4.1. Trust me

Trust isn't just a solution—it's a natural, intuitive one. That may look doubtful. But I wouldn't waste your time (or mine) with a gerrymandered formal principle. Right off the bat, there are four things you need to know.

One: Trust is symmetric. Say that a principle **holds** at a world w iff all of its instances (well-defined instantiations of free variables) are true at w . Then:

Fact 4.1: *In any probability frame: Trust holds at a world iff $P_q(p|P_q(p) \leq t) \leq t$ does.*

So upon learning that the expert judges that p ($P(p) \geq t$), you should judge that p ; and upon learning that the expert *doesn't* judge that p ($P(p) \leq t$), you *shouldn't* judge that p .

Uh oh. Does this imply Reflection? I say that (Trust:) upon learning that the expert is at least 0.7 confident, be at least 0.7; and upon learning that they're at most 0.8 confident, be at most 0.8. Does it follow that (Reflection:) upon learning that the expert is at least 0.7 *and* at most 0.8 confident, you should be at least 0.7 and at most 0.8? No.

Two: Trust does not imply Reflection.

Fact 4.2: *There are probability frames that validate Trust in which Reflection fails at all worlds.*

This is possible because probabilistic support is non-monotonic: learning one thing can push your credence above 0.7, even if further learning would pull it lower. And it is actual because of the crux of our story: an expert judgment is never evidence against p , while an expert opinion sometimes is. Example: recall Immodest Imani, who is certain that Imani's the expert (I). If you learn that the expert judges that Imani's the expert to degree 0.7 ($P(I) \geq 0.7$), you should raise your credence to at least 0.7—for you know the expert would react to this information by doing so. But if you further learn that the expert both judges that Imani's the expert to degree 0.7 *and* does *not* judge that she is to degree 0.8 ($[P(I) \geq 0.7] \wedge [P(I) < 0.8]$), you should drop your credence to 0—for you know that Imani *does* judge that she's the expert to degree 0.8. In short, Trust works because—and only because—it applies to judgments. It does not imply Reflection.

Three: Trust implies special cases of Reflection. Suppose you should be certain that the expert's opinion is in a given range: $P(P(p) \in [l, h]) = 1$. Then—since updating on these bounds doesn't provide any new information—Trust constrains your *unconditional* credences to be in that range.¹⁹ More generally, when you should be sure that *if* the evidence warrants a given opinion, *then* it warrants certainty that it does so, you should obey Reflection. Letting \mathbf{Sp} ($\text{=}_{df} P(p) = 1$) mean that your evidence warrants being Sure of p :

Fact 4.3: *In any probability frame, if Trust holds, and for a given $l, h \in [0, 1]$ it's true that $P(S(P(p) \in [l, h])|P(p) \in [l, h]) = 1$, then $P(p|P(p) \in [l, h]) \in [l, h]$.*

¹⁹ Precisely, $l \leq P(p|P(p) \geq l) = P(p) = P(p|P(p) \leq h) \leq h$. This is a synchronic version of a "conglomerability" constraint (Easwaran 2013). The generalization of Trust in §7 implies a more standard diachronic version: the current rational credence is bounded by the possible future rational credences.

Four: paraphrased into natural language, Trust is truistic. That may seem doubtful, for Trust governs only judgments—and even if they are theoretically important, judgments may seem a contrivance.

They are not. Our natural-language talk of confidence trades in judgments. We talk about being $\left| \begin{array}{c} \text{very} \\ \text{fairly} \\ \text{sorta} \end{array} \right|$ *confident* of *Top*, about it being $\left| \begin{array}{c} \text{really} \\ \text{pretty} \\ \text{somewhat} \end{array} \right|$ *likely* to be true, about

$\left| \begin{array}{c} \text{leaving open} \\ \text{suspecting} \\ \text{thinking} \end{array} \right|$ that it's true, and so on. All of these terms share an important logical fea-

ture: they are preserved under increases in probability. If you are *sorta* confident of *Top* and then you become more confident, you are *still* *sorta* confident of it (though you may now also be very confident); if it's *somewhat* likely to be true and then becomes more likely, it is *still* *somewhat* likely to be true (though it may now also be really likely); if you leave open that it's true and then you become more confident, you *still* leave open that it's true (though you may now also think that it is).²⁰ It follows that these terms cannot denote (proper) opinions, for such opinions are bounded above and below—if you have a middling opinion that *Top*, then increasing your confidence can lead you to lose it. Instead, these terms must denote *judgments*. Combining this observation with the fact that natural language expresses conditional probabilities as probabilities of (indicative) conditionals (cf. Stalnaker 1970; Adams 1975; Edgington 1995), we see that Trust is really just a set of truisms:

You should think it $\left| \begin{array}{c} \text{really} \\ \text{pretty} \\ \text{somewhat} \end{array} \right|$ likely that if the evidence makes it $\left| \begin{array}{c} \text{really} \\ \text{pretty} \\ \text{somewhat} \end{array} \right|$ likely that *p*, then *p*.

You should be $\left| \begin{array}{c} \text{very} \\ \text{fairly} \\ \text{sorta} \end{array} \right|$ confident that if the evidence warrants being $\left| \begin{array}{c} \text{very} \\ \text{fairly} \\ \text{sorta} \end{array} \right|$ confident that *p*, then *p*.

You should $\left| \begin{array}{c} \text{think} \\ \text{suspect} \\ \text{leave open} \end{array} \right|$ that if the evidence warrants $\left| \begin{array}{c} \text{thinking} \\ \text{suspecting} \\ \text{leaving open} \end{array} \right|$ that *p*, then *p*.

And if Lockeanism is right that belief reduces to sufficiently high credence (Foley 1992, 2009; Sturgeon 2008; Leitgeb 2013; Dorst 2017):

You should believe that if the evidence warrants believing that *p*, then *p*.

Upshot: Trust is an elegant, well-motivated weakening of Reflection.

4.2. Trust Trust

The progression of this paper suggests two hypotheses. The first:

Trust vindicates our Modest and Guiding Truisms.

I claim that I can establish this first hypothesis—that Bridging succeeds, after all. In fact, this success is so resounding that it suggests that Splitters and Mergers should jump ship.

²⁰ Saying “It’s rather likely to rain” suggests that it’s *not* very likely to. But this is a pragmatic phenomenon—compare: “Jane is rather tall” implicates (but does not entail) that she’s not very tall.

Precisely, say that you are *higher-order coherent* iff there is a body of evidence that would make your first- and higher-order opinions rational. The second hypothesis:

You are higher-order coherent iff you obey Trust.

The rest of this paper defends these two hypotheses. §5 shows that Trust both vindicates our Modest Truism and has many guiding features. However, it also has commitments—most notably, the *positive access* principle that if you should be sure of p , you should be sure that you should be. In light of this, some may seek a weaker principle. §6 replies that we *need* a strong principle like Trust to avoid paradoxical results. §7 goes further: an independent characterization of the Guiding Truism leads *exactly* to Trust. §8 closes with applications.

It is the results in §5 and §7 that establish the first hypothesis. And it is the entire story—from the necessity of Bridging, to the failures of alternatives, to the successes of Trust—that composes the argument for the second.

5. Of Trust

In this section we'll see the virtues of Trust: it allows you to be extremely modest, while still requiring that you treat your evidence as a guide. Here's what you need to know. On the modesty side: Trust allows you to be certain that the expert is modest, as well as to be virtually indifferent over which of an arbitrary set of opinions is rational. On the guiding side: Trust requires you to react to new information as you know the expert would; it requires that you think the expert's judgment is correlated with truth; it rules out our puzzling case of Misguided Evidence; and it requires that the more higher-order doubts you have, the more moderate your first-order opinions must be. This is all exactly what we want. But Trust also has controversial commitments: it implies the *surely-factivity* principle that you should be sure that your evidence only warrants being certain of p if p is true; and it implies the *positive access* principle that if you should be sure of p , you should be sure that you should be sure of p . We might worry about these commitments—but §§6–7 argue that we *need* them. For those mainly interested in the big picture, the rest of §5 can be skipped. For those interested in the details, follow me into the weeds.

Modesty first. Trust isn't just formally weaker than Reflection—it's *substantively* weaker. First, Trust allows (but does not require!) you to be certain that your evidence warrants modesty. Letting *modest* be the proposition that you should be uncertain what the rational credence function is:

Fact 5.1: *There are probability frames that validate both Trust and $[P(\text{modest}) = 1]$.*

Of course, Fact 5.1 doesn't tell us *how* modest Trust allows you to be. But it turns out that Trust allows you to be extremely modest—you can be virtually indifferent as to which of an arbitrary set of opinions is rational:

Fact 5.2: *For any $T = \{t_1, \dots, t_n\} \subset [0,1]$, there are probability frames that validate Trust with a candidate²¹ π and proposition p such that $\pi(P(p) \in T) \approx 1$ and for all t_i : $\pi(P(p) = t_i) \approx \frac{1}{n}$.*

²¹ Unlike ' P ', ' π ' is a rigid designator for a probability function whose values are known. π is a **candidate** in a probability frame iff at some world you should leave open that π is the rational credence function.

Upshot: our diagnosis of Reflection was correct. The *reason* it ruled out modesty was that it failed to distinguish facts about evidential support that can and cannot be evidence against p . Trust is built upon this distinction—and as a result, it vindicates our Modest Truism.

What about our Guiding Truism? There are two separable components to being properly guided. Start with a metaphor. If you're going on a hike, you need two things—a guidebook and a compass. The guidebook has no information about your immediate surroundings. It's useful because it has a wealth of *conditional* information—"If you're on the path facing east, the road is to your left", and so on. The guidebook is valuable as a *map*—once you enrich it with your local knowledge, it gives fine-grained directions; it is an optimal handler of new information. In contrast, the compass can't give you fine-grained directions. Its purpose is simply to orient you toward something you care about—in this case, north. The compass is valuable as a *tracker*; it is a reliable indicator of north. Having a map but no tracker is liable to lead you astray—if you're turned around, the map will send you in exactly the wrong direction. Having a tracker but no map is liable to leave you in the dark—if you don't know the area, you'll walk right past your campsite. To be properly guided, you need *both* a map and a tracker.

End of metaphor. For your *evidence* to be a guide, it must have the features of both a map and a tracker. More precisely, it must be both (1) an *optimal handler of new information* and (2) a *reliable indicator of the truth*.

(1) First, an optimal handler of new information—a map. When the expert knows everything you know, you should defer to their opinions. This doesn't require Reflection—for when you learn about the expert's opinion you may know something that they don't. (Looking at your surroundings, you may know more than your map does.) Rather, what it requires is that whenever you get new information, you should react to it *as you know the expert would*. Imagine the following scenario. Sitting on the panel, everyone hears an announcement that q . (q could tell you anything—e.g. that Bill is not the expert.) After q is announced, the remaining panelists—the ones who still might be the expert—announce their new opinions in *Top*. How, then, should *you* react to q ? Slogan: *react as you know the expert would*. If all remaining panelists react with opinions about *Top* in a given range, then *you* should have that opinion in that range. Precisely²²:

Reaction: If $P_q(P_q(p) \in [l, h]) = 1$, then $P_q(p) \in [l, h]$

If you should be sure that your evidence warrants reacting to your information q with a given opinion, react with that opinion.

Slogan: *react as you know the expert would*.

Reaction encodes an important component of treating your evidence as a guide, for it guarantees that your opinions are constrained by your opinions about the expert's opinions. But—as will become clear in §6—just like a map without a compass, Reaction by itself would allow your evidence to send you in exactly the wrong direction.

(2) So evidence must also be a *tracker*—the expert's best guesses should be a reliable indicator of the truth. (The pointer on the compass should be a reliable indicator of north.) Imagine the following scenario. Everyone on your panel is asked whether or not

²² Reaction is inspired by Elga's (2013) "New Reflection" principle (cf. Hall 1994), but Reaction is stronger. Moreover—unlike New Reflection (cf. Pettigrew and Titelbaum 2014)—Reaction is preserved under conditioning.

they judge that *Top* (i.e. whether or not their credence is above a salient threshold). You are then going to find out what the expert did: you'll see a green light if the expert judged that *Top*, a red light otherwise. You have some prior opinion about *Top*. Then you see a green light—how should your opinion change? You now know that the expert judged that *Top* is true. If expert judgments track the truth—if their confidence in *Top* is correlated with its truth value—this is evidence in favor of *Top*. So you shouldn't decrease your confidence in *Top*. Precisely, and generalizing to conditional judgments:

Reliance: $P_q(p|P_q(p) \geq t) \geq P_q(p)$

Upon learning that your evidence warrants reacting to your information q by judging that p , you shouldn't decrease your confidence in p .

Slogan: *take experts to be reliable*.

We have two components to being a guide: being an optimal responder to new information (a map) and being a reliable indicator of the truth (a tracker). Trust vindicates both. In fact, Trust vindicates *exactly* both:

Fact 5.3: *A probability frame validates Trust iff it validates both Reaction and Reliance.*

This is the formalization of the intuitive argument from §4: if we require that you react to new information as you know the expert would and that expert credences are correlated with truth, then Trust is precisely what we get. Upshot: Trust yields the general guiding features—of a map and a tracker—that we're after.

It also yields specific ones. First, Trust rules out our paradigm cases of a mismatch between first- and higher-order opinions. Recall Misguided Evidence: putative cases in which it's rational to have confidence that p but I shouldn't have confidence that p ($\exists p, t : P(p \wedge [P(p) < t]) \geq t$). Trust prevents this:

Fact 5.4: *If Trust holds at a world in a probability frame, Misguided Evidence does not.*

Finally, Trust forces a robust bridge between your first- and higher-order attitudes:

Fact 5.5: *In any probability frame: if Trust holds then $[P(P(p) \geq t) \geq s] \rightarrow [P(p) \geq t \cdot s]$ does too. No stronger connection holds: for any $t, s \in [0, 1]$: there are probability frames that validate Trust and make both $[P(P(p) \geq t) \geq s]$ and $[P(p) = t \cdot s]$ true at a world.*

Fact 5.5 implies that if you're confident that you should be confident of something, you should be at least somewhat confident of it. Example: if you should be at least 0.9 confident that you should be at least 0.9 confident of *Top*, then you should be at least 0.81 confident of *Top*. Conversely, if you should have significant higher-order doubts, your first-order opinion must be moderate. Example: if you should be at least 0.4 confident that you should be at least 0.6 confident ($P(P(p) \geq 0.6) \geq 0.4$) but *also* at least 0.4 confident that you should be *at most* 0.4 confident ($P(P(p) \leq 0.4) \geq 0.4$), then your opinion in p must be between 0.24 and 0.76. More generally, if you are to have an opinionated credence in p , you can't have many higher-order doubts. For instance, if your credence in p should be 0.1 ($P(p) = 0.1$), then: the maximal credence you can have that you should be at least 0.2 confident ($\max[P(P(p) \geq 0.2)])$ is $\frac{1}{2}$; the maximal credence you

can have that you should be at least 0.3 confident is $\frac{1}{3}$; ... and the maximal credence you can have that you should be at least 0.9 confident is $\frac{1}{9}$. In contrast: if you have a moderate credence in p , your higher-order doubts can range much wider. For instance, if your credence in p should be 0.5 ($P(p) = 0.5$), then: the maximal credence you can have that your credence in p should be at least 0.6 is $\frac{5}{6}$; the maximal credence you can have that your credence in p should be at least 0.7 is $\frac{5}{7}$; ... and the maximal credence you can have that your credence in p should be at least 0.9 is $\frac{5}{9}$. Using Fact 5.5, these relationships are graphed for various values of $P(p)$ in Figure 1 below.²³

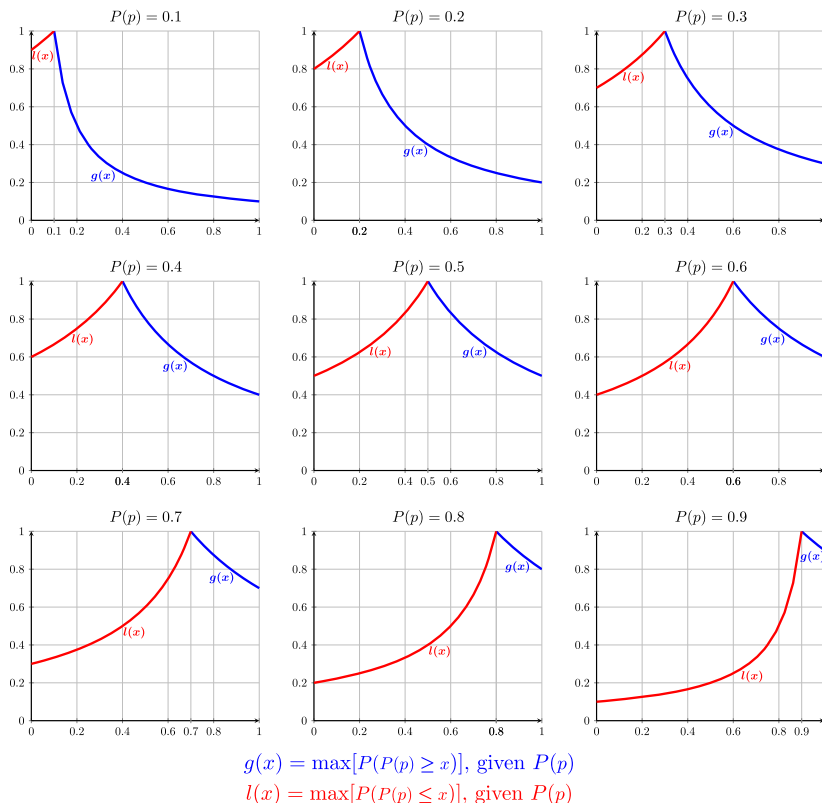


Fig. 1. Possible higher-order doubts, given first-order credence. [Color figure can be viewed at wileyonlinelibrary.com]

So we know a lot about Trust. How? The key to many of the results stated in this section is that we can give an *exact* characterization of Trust within a natural class of models (Theorem 5.7, Appendix A.1).²⁴ The upshot is that Trust imposes three substantive—but tenable—constraints on the structure of evidence.

The first constraint is *surely-factivity*. You should be sure that: you should be sure of p only if p is true. Recalling that Sp is the proposition that your evidence warrants being

²³ The graphs do not plot what your higher-order opinions *must* be, given your first-order ones—they represent how Trust *constrains* the relation between the two. Note: the maxima are not always jointly satisfiable: if $P(p) = 0.1$ you (obviously) cannot have both $P(P(p) \geq 0.1) = 1$ and $P(P(p) \leq 0) = 0.9$.

²⁴ Those models are *prior frames*—a subclass of probability frames in which uncertainty about what you should think flows from uncertainty about what you should be sure of (i.e. condition on).

Sure of p : $S(Sp \rightarrow p)$. In other words, you should never say to yourself, “Maybe I should be certain of p even though it’s false” (i.e. $P(Sp \wedge \neg p) > 0$). (For if you did, then upon learning that your evidence warrants certainty of p , you’d still be less than certain of it—you wouldn’t trust your evidence.) Surely-factivity makes full-blown factivity (i.e. $Sp \rightarrow p$) natural, though not inevitable.

The second constraint is *positive access*. You should be sure of p only if you should be sure that you should be: $Sp \rightarrow SSp$.²⁵ In other words, you should never say to yourself, “ p is true, but maybe I shouldn’t be sure of it.” (i.e. $Sp \wedge [P(\neg Sp) > 0]$). (For if you did, then upon learning that your evidence warrants being *unsure* of p , you’d still be sure of it—you wouldn’t trust your evidence.)

Surely-factivity and positive access are the only non-probabilistic constraints that Trust imposes in full generality. But under the auxiliary assumption that your higher-order uncertainty comes from uncertainty about what you should be sure of, Trust is closely related to a third constraint. It is implied by (and *almost* implies) the constraint that we can define an indicative conditional operator ‘ \rightarrow ’ that interacts sensibly with your conditional beliefs: in particular, (a) if you are certain of p conditional on q , then you are certain of *if q then p* ; and (b) for consistent q , if you are certain that *if q then p* then you are not certain that *if q then $\neg p$* . (See Appendix A.1 for details.)

Here, then, is where we are. Trust is the culmination of our search for a modest guide: it is a well-motivated refinement of Reflection, it allows plenty of modesty, it captures both the *map* and *tracker* aspects of being a guide, it rules out our puzzle cases, and it forces a sensible connection between your first- and higher-order attitudes. In short, we have reason to think that Trust is the Goldilocks principle.

But not conclusive reason. Trust imposes controversial constraints on the structure of evidence—most notably, positive access. Many will want to reject this principle—and will therefore be worried about Trust. Bridgers may try to endorse a weaker principle (cf. Elga 2013; Sliwa and Horowitz 2015; Christensen 2016), while Splitters may reject the search for such principles altogether (cf. Lasonen-Aarnio 2015; Williamson 2018). What of such alternative approaches—might they succeed? No.

6. In Judgment

This section argues that without a strong principle like Trust, paradox ensues. Though the argument is forceful, it is blunt—it cannot pinpoint Trust as the only solution. But a follow-up argument can: §7 proposes a characterization of our Guiding Truism in terms of the *value of evidence* (Good 1967), and then shows that it leads *exactly* to Trust.

Why won’t a weaker principle do? Because there is another puzzle—one which, if allowed, would undermine the normative role of evidence. Consider:

SYCOPHANTS

Looking at the spoons, Sybil and Phan should be sure that one of them—the diligent one—has the credences warranted by their (shared) evidence. Conditional on it being either one of them, they agree on everything. But unconditionally they disagree: Sybil is 0.9 confident that Phan is the diligent one; Phan is 0.9 confident that Sybil is.

²⁵ Trust does *not* require *negative access*, which says that if you shouldn’t be sure of p , you should be sure that you shouldn’t be: $\neg Sp \rightarrow S\neg Sp$. Unlike its positive counterpart, negative access rules out modesty (within prior frames) and is untenable for factive attitudes like knowledge (Stalnaker 2006).

They are falling over themselves to give each other credit:

- Sybil: ‘You’re probably more diligent than me.’
 Phan: ‘No, *you’re* probably more diligent than *me*.’
 Sybil: ‘No, **YOU’RE** probably more diligent than **ME**.’
 Phan: ‘No!’ [...]

Schematically, the scenario looks like this. There are two relevant possibilities: that Sybil is the diligent one (s) and that Phan is (p). What they should think depends on which of these possibilities they’re in. If Sybil is the diligent one, they should be 0.9 confident that Phan is the diligent one and (so) 0.1 confident that Sybil is. If Phan is the diligent one, they should be 0.9 confident that Sybil is the diligent one and (so) 0.1 confident that Phan is. Letting an arrow labeled t from x to y mean that at possibility x they should assign credence t to being in possibility y , we have Figure 2.

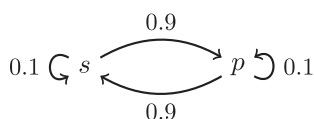


Fig. 2. Sycophants.

We shouldn’t allow Sycophants: doing so would thereby allow you to be certain that your evidence is an anti-guide—that the expert has made an error in judgment. p is the proposition that Phan is the diligent one. If p is true, then their evidence warrants judging—as Phan does—that that p is false: $P(\neg p) \geq 0.9$. And if p is false, then their evidence warrants judging—as Sybil does—that p is true: $P(p) \geq 0.9$. Either way, their evidence warrants making the wrong judgment about p : at both possibilities the biconditionals $p \leftrightarrow [P(\neg p) \geq 0.9]$ and $\neg p \leftrightarrow [P(p) \geq 0.9]$ are true. Since Sybil and Phan should be sure that one of these possibilities is actual, they should be *sure* of these biconditionals. Thus they should be sure that *Phan is the diligent one iff we should judge that he’s not*, and that *Phan is not the diligent one iff we should judge that he is*: $S(p \leftrightarrow [P(\neg p) \geq 0.9])$ and $S(\neg p \leftrightarrow [P(p) \geq 0.9])$. Generalizing, the puzzle is that you should be sure that the rational credence in p is anti-correlated with the truth:

Effacing Evidence: $\exists p, t : S(p \leftrightarrow [P(p) < t])$ and $S(\neg p \leftrightarrow [P(p) > t])$.

You should be sure that p is true iff you should not have confidence in it, and that p is false iff you should have confidence in it.

I claim that any theory that allows Effacing Evidence is too weak.

First, intuitively. Suppose we are trying to figure out what Phan thinks:

- Phan: ‘I’ve judged that Sybil is the diligent one.’
 Us: ‘What does *she* think?’
 Phan: ‘She’s judged that I’m the diligent one. So I probably ought to judge that I am.’
 Us: ‘Well, why don’t you?’
 Phan: ‘I said I probably *ought* to judge that I’m the diligent one—not that doing so would probably be correct. It would most likely be incorrect.’
 Us: ‘But why do you think that it would most likely be incorrect? Is that what your evidence suggests?’

- Phan: ‘To the contrary. My evidence probably suggests that I’m the diligent one. That’s why I think that Sybil is.’
- Us: ‘Wait—you’re not listening to your evidence?’
- Phan: ‘I sure hope not! My evidence is wrong about who the diligent one is.’
- Us: ‘So why conform to your evidence?’
- Phan: ‘Exactly. Of course, I *ought* to—my evidence determines what I should think, after all. But it’s a bad idea this time around. That’s why I’m trying to avoid it.’
- Us: ‘No: why *ought* you conform to your evidence, if you know it’ll lead you astray?’
- Phan: ‘... I just ought to! My evidence tells me what I ought to think. Sometimes I know that what it tells me to think is false. But what am I to do? Life is hard.’

Something has gone wrong. A theory of rationality can’t be made of just any arbitrary list of requirements—it must have the right structural features. In particular, the theory must be able to answer the question: ‘Why conform to my evidence?’ (Horowitz 2013; Schoenfield 2015b). No theory that allows Effacing Evidence can do so. For—as Phan’s thoughts reveal—such theories sometimes allow you to know that your evidence is an anti-guide to what you actually value.

We can make this precise in two ways. Suppose you have Effacing Evidence with $t = \frac{1}{2}$: you should be sure that p is true iff you should be less than $\frac{1}{2}$ confident of it, and that p is false iff you should be more than $\frac{1}{2}$ confident of it. (The arguments generalize.)

First, you can be certain that choosing rationally will lead you to lose money. For I can offer you the following three options:

- Bet 1: win \$1 if p , lose \$1 if $\neg p$
- Bet 2: win \$1 if $\neg p$, lose \$1 if p
- No Bet: \$0

It’s rational (maximizes expected utility) to take Bet 1 iff you should be *more* than $\frac{1}{2}$ confident of p . It’s rational to take Bet 2 iff you should be *less* than $\frac{1}{2}$ confident of p . But you should be certain that you should be more than $\frac{1}{2}$ confident of p (should take Bet 1) iff $\neg p$ —i.e. iff you will lose that bet. And you should be certain that you should be less than $\frac{1}{2}$ confident of p (should take Bet 2) iff p —i.e. iff you will lose *that* bet. So you should be certain that acting rationally—taking the option you *ought* to take—will lead you to lose \$1 and be poorer than if you took No Bet! Upshot: theories that allow Effacing Evidence allow rational requirements to lead to a sure loss.²⁶

Second, you can be certain that being rational will lead you to lose accuracy. For one of your doxastic options is to have credence $\frac{1}{2}$ in p , come what may. Yet you should be certain that it’s rational to have credence *below* $\frac{1}{2}$ in p iff p is true, and that it’s rational to have credence *above* $\frac{1}{2}$ in p iff p is false. That is, being rational will either lead you to have less confidence in a truth or more confidence in a falsehood than simply having credence $\frac{1}{2}$. So you should be certain that believing rationally—having the credence you *ought* to have—will lead you to be less accurate than having credence $\frac{1}{2}$. Upshot: theories

²⁶ Many take it as a premise that rational requirements can do no such thing—most notably, Dutch Bookies (cf. Ramsey 2010; Skyrms 1966; Christensen 1991).

that allow Effacing Evidence allow rational requirements to be accuracy-dominated by another available option.²⁷

In short, theories that allow Effacing Evidence allow you to be certain that being rational—in and of itself—will prevent you from getting the things you actually value, like money and accuracy. They cannot vindicate our Guiding Truism.

But wait—is it really so surprising that sometimes you can know that being rational will make you worse off? Suppose I credibly threaten to drain your bank account iff your credences are rational. You then can be certain that being rational will prevent you from keeping the things you actually value (like money), right?

Right. But in this case there is a *cost* to being rational: doing so puts you in a bad situation where—no matter what else you do—you will be left with less money. Not so in my argument against Sycophants. In that argument there is absolutely no cost to being rational: all the same options have all the same values, regardless of whether you are rational or not. (If you are rational, it is still true that *were you to take No Bet, you wouldn't lose any money.*) The only reason you can be certain that being rational will lead you to lose money is that you can be certain that being rational will lead you to choose the *worst* available option. The difference is crucial. All theories need to allow that sometimes you can know that being rational will lead you to be hurt. But no theory should allow that sometimes you can know that being rational will lead you to *hurt yourself*.

But wait—*couldn't* you have Effacing Evidence? Suppose God sets things up so that the rational opinions are inaccurate about *p*—and then she announces as much. Can you now be sure that being rational will lead you to have inaccurate opinions about *p*? No. Distinguish two versions of the case.

First version: God sets things up so that you are rational to be confident of *p* iff *p* is false: $[P(p) \geq 0.7] \leftrightarrow \neg p$. Then she announces this fact, so you update your beliefs on the claim that $[P(p) \geq 0.7] \leftrightarrow \neg p$. Do you now have Effacing Evidence? No. For what you've learned was that *before God told you anything*, the rational credence in *p* was pointing in the wrong direction. This does not imply that *after* being told this, the updated rational credence is still pointing in the wrong direction. Formally, letting $q =_{df} ([P(p) \geq 0.7] \leftrightarrow \neg p)$, it is true that $P_q([P(p) \geq 0.7] \leftrightarrow \neg p) = 1$ but it does not follow (and is not plausible) that $P_q([P_q(p) \geq 0.7] \leftrightarrow \neg p) = 1$, as required for Effacing Evidence.

Second version: God offers the self-referential announcement, “The rational reaction to this very announcement is to be certain that: the rational credence in *p* is high iff *p* is false.” In response to this announcement, do you have Effacing Evidence? No. For note that a parallel objection can be given against virtually any other rational norm—God could just as well announce: “The rational reaction to this very announcement is to be certain of $p \wedge \neg p$.” Since this “counterexample” to the norm not to believe contradictions is not compelling, neither is the “counterexample” to a ban on Effacing Evidence.

I conclude, then, that we should reject theories that permit Effacing Evidence. What theories do so? Many. I cannot go into the details here, so a summary will have to do. First, the Sycophants model validates positive access ($Sp \rightarrow SSp$), negative access ($\neg Sp \rightarrow S\neg Sp$), and factivity ($Sp \rightarrow p$), so traditional versions of access internalism cannot rule it

²⁷ Many take it as a premise that rational requirements can do no such thing—most notably, epistemic utility theorists (e.g. Joyce 1998, 2009; Pettigrew 2013, 2016).

out (e.g. Smithies 2012). Second, since Sybil and Phan agree on everything upon learning which one of them is rational, the Sycophants model validates Reaction.²⁸ Finally, evidential versions of calibrationism (Sliwa and Horowitz 2015; Christensen 2016) can generate cases of Effacing Evidence, as can natural versions of Williamson's (2014) unmarked clock model (see Dorst 2018b). If I am right, no such theory can be correct. Instead, we need a principle strong enough to rule out Effacing Evidence. Trust fits the bill:

Fact 6.1: *If Trust holds at a world in a probability frame, Effacing Evidence does not.*

Upshot: despite its controversial consequences, we *need* a strong theory like Trust.

But how can we be sure that we've got that theory just right? What if Trust is too weak—permitting as-yet-unnamed puzzles? Or what if Trust is too strong—ruling out more than is required to solve them? I'm now going to offer a *proof*—of sorts—that it's not. No stronger theory is needed, and no weaker theory will do.

7. Of Value

What we need is *an* explication of our Guiding Truism that comes out true. I'm going to defend a proposal for the *minimal* such explication—the weakest principle that fully captures our Guiding Truism—and then show that it leads *exactly* to Trust. Thus Trust vindicates *an* explication—no stronger theory is needed. And Trust vindicates the *minimal* explication—no weaker theory will do.

The idea is this. Why should we care about gathering and conforming to evidence? What makes evidence something of value? I.J. Good (1967) gave a famous answer: evidence helps us make good choices. Suppose you face a decision problem—a set of options that may lead to different outcomes to which you assign different values. A body of evidence is *valuable* iff—supposing the evidence is free—the expected value of taking the option warranted by the evidence (whatever it is) is higher than the expected value of any other particular option. (Loosely: iff you should prefer to *use* free evidence to make your decision, rather than ignore it.) Example: you have to decide whether to take a bet on *Top*. You could either (1) scrutinize the spoons more closely (obtain more evidence) or (2) simply decide now whether to take the bet (ignore that evidence). No doubt the expected value of (1) is higher than that of (2)—the evidence is valuable.

Good generalized that idea. He argued that—no matter what options and values you have—evidence is *always* valuable. In formulating this idea, many have treated it diachronically: you should prefer to gather and use a *more* informed body of evidence to make your decision (cf. Skyrms 1990; Oddie 1997; Myrvold 2012; Huttegger 2014; Ahmed and Salow 2018). But once we allow modesty—uncertainty about your *current* evidence—there are in fact two instances of Good's idea. In making your decision, you should prefer: (1) to use your current evidence (whatever it is), rather than simply choose an option; and (2) to use a more informed body of evidence (whatever it is), rather than simply choose an option.

We can capture both instances under one schema. Say that a body of evidence is *at least as informed as yours* (for short: **informed**) iff it contains all your evidence, and

²⁸ Hence it also validates Elga's (2013) New Reflection principle, mentioned in footnote 22.

maybe more. So (trivially) your evidence is at least as informed as your evidence, and (nontrivially) the evidence you'd have after scrutinizing the spoons is at least as informed as your evidence. Generalizing Good's idea (formalized in Appendix A.2):

Value

You should always expect the option warranted by informed evidence to be at least as good as you should expect any other particular option to be.

Slogan: *evidence is valuable*.

Example: I offer you a bet on *Top*—you must decide whether to take it or leave it. If you were 0.8 confident of *Top*, you'd take it; if you were 0.6 confident of *Top*, you'd leave it. Being 0.7, you're on the fence—the expected value of taking it and of leaving it are balanced. What about the expected value of doing what you *should* do (what the expert would do)—whatever that is? If you should be 0.8 (as perhaps you should), what you should do is take it; if you should be 0.6 (as perhaps you should), what you should do is leave it. Being unsure what you should think, you're unsure what you should do. But why care about doing what you *should* do? Why treat your evidence as a guide? Value says: because you should expect it to help you make good decisions.

Value applies to any informed body of evidence—not just your own. Following suit, we can generalize our principles; for example, our Guiding Truism:

INFORMED GUIDING TRUISM

Your evidence always warrants being disposed to treat *informed* evidence as a guide.

Similarly for our deference principles. Let i and k be two (perhaps identical) bodies of evidence—subject to the constraint that k is at least as informed as i ($k \geq i$). Let P^i be the credences that are rational given evidence i (whatever they are), and P^k be those that are rational given evidence k . Then Trust (and its ilk) should be generalized:

Informed Trust: $P_q^i(p|P_q^k(p) \geq t) \geq t$ ($k \geq i$)

Upon learning that *informed* evidence warrants reacting to q by judging that p , react to q by judging that p .

Slogan: *take informed expert judgments on trust*.

Having generalized our search for a modest guide, I'll now argue that Value is the *minimal explication* of the (Informed) Guiding Truism—it fully vindicates that truism; and no weaker principle would.

As with the argument against Effacing Evidence in §6, it is crucial to note that Value applies to being rational *in and of itself*. Three clarifications. (1) Value applies to the expected value of *freely* taking the option warranted by an informed body of evidence; thus it screens off any costs—monetary, psychological, computational, and so on—that using that evidence may have. (If you will be punished for using your evidence, doing so might not maximize expected value.) (2) Value applies to the expected value of *successfully* taking the option warranted by an informed body evidence; thus it screens off any risks of *misusing* the evidence that obtaining it may bring. (If trying to use the evidence will likely lead to mistakes, doing so might not maximize expected value.) (3) Value applies to the expected value of *simply* taking the option warranted by an informed body of evidence; thus it screens off any effects on you—such as changes in values—that obtaining that evidence may have. (If using the evidence will change your outlook on life (Paul 2014), doing so might not maximize expected value.) Although

such costs, risks, and effects are ever-present in real life, Value explains why—when they become sufficiently small—you should always prefer to use the evidence.

Suitably clarified, Value is plausible. It is also applicable—it explains what’s puzzling about our puzzles:

Fact 7.1: *In any dynamic probability frame: if Misguided Evidence or Effacing Evidence are true at a world, Value fails.*

This is because Value rules out the possibility that evidence could be an “anti-guide,” and thus subsumes the argument against Effacing Evidence in §6. (A similar argument can be given against Misguided Evidence.)

Finally, here is a general argument that Value is the minimal explication of the (Informed) Guiding Truism—the claim that you should always be disposed to treat informed evidence as a guide. There is no doubt that it is *an* explication. For suppose that Value holds. Then no matter what decision problem you’re facing—whether it’s what to do, or what to think—then if you *could* do what’s warranted by an informed body of evidence, then you *should*. (Doing so would maximize expected value). Thus there’s a perfectly good sense in which you should always treat informed evidence as a guide—our Guiding Truism comes out true. Conversely, no principle *weaker* than Value would vindicate our Guiding Truism. For suppose that Value fails: there is a decision problem in which you expect that doing what’s warranted by an informed body of evidence will lead to a worse outcome than ignoring that evidence. So if you were given a choice between the two, you should prefer to ignore the evidence—for you should expect the expert (the person following the evidence) to make a worse choice than you! Thus there’s *no* perfectly good sense in which you should *always* treat informed evidence as a guide—our Guiding Truism comes out false.

Suppose that this is right: we vindicate our Guiding Truism iff we vindicate Value. What does it take to vindicate Value? Famously, Good (1967) proved it under certain assumptions. But it turns out that those assumptions were inconsistent with our Modest Truism.²⁹ What happens when we loosen them?

Recall our story. In searching for a modest guide, we discovered that we must: permit modesty; forbid Misguided Evidence; permit Reflection failures; require Reaction; require Reliance; require surely-factivity; require positive access; forbid Effacing Evidence; require Trust.

Value runs the gamut. It permits modesty. It forbids Misguided Evidence. It permits Reflection failures. It requires Reaction. It requires Reliance. It requires surely-factivity. It requires positive access. It forbids Effacing Evidence. It requires Trust:

Theorem 7.2: *In any dynamic probability frame: if Value holds at a world, Informed Trust does as well.*

In fact—in (at least) a wide class of scenarios—it is *equivalent* to Trust:

Theorem 7.4 (Rough): *In (at least) a wide class of frames: Value \Leftrightarrow Informed Trust.*

²⁹ Precisely, he assumed that we could model you using a prior frame in which E is an equivalence relation (see Appendix A.1). Any such frame validates $[P(p) = t] \leftrightarrow S[P(p) = t]$.

That wide class is the class of *prior* frames discussed in §5 and Appendix A: models in which your higher-order uncertainty stems from uncertainty about what you should be sure of. Moreover, the restriction is on the proof—not the truth. There is reason to think that (Informed) Trust is equivalent to Value in full generality. I conjecture that it is (Conjecture 7.3, Appendix A.2). But even if that conjecture fails, we now know that some *strengthening* of Trust will succeed. Upshot: we *know* that Trust is necessary—and in at least a wide class of cases, sufficient—to vindicate our Guiding Truism. And we have reason to think that Trust *characterizes* our Guiding Truism: that no stronger theory is needed, and no weaker theory will do.

A final point: Theorem 7.4 is a “coincidence result.” The progression of this project was not so prescient as the progression of this paper: I began with various puzzles of higher-order uncertainty, was led (through trial and error) to Trust, and characterized it over a class of prior frames. Only later did I discover that Geanakoplos (1989) had proven that (a principle like) Value was validated by *the exact same class of prior frames*. Imagine my surprise—shock even—upon seeing his theorem. And my satisfaction upon discovering it could be strengthened, yielding our coincidence. Personally, I think this stroke of serendipity increases the plausibility of our destination. Perhaps you will agree.

This is the end of our story. Wanting to take disagreement seriously, we began searching for a modest guide. The obvious proposal—Reflection (*defer to expert opinions*)—was too strong, for it failed to acknowledge that learning an expert *opinion* can be evidence against *p*. The natural response—Trust (*take expert judgments on trust*)—had all the marks of the Goldilocks principle. And the measure of success—Value (*evidence is valuable*)—gave Trust a resounding confirmation. Our search for a modest guide has succeeded: those who are uncertain can take it on trust.

8. In Consequence

I’ll close by sketching two applications of the theory developed here. Both deserve more discussion; my goal is just to sketch how Trust is rich in consequences for other debates.

8.1. To Disagreement

We began with a disagreement: you were confident of *Top*, while Disagreeing Disa was not. Intuitively, learning that she was not should lead you to lower your confidence in *Top*. And we saw that if this is to be so, then our Modest and Guiding Truisms must both be true. Having found a theory that reconciles them, we can vindicate the intuitive verdict.

You and Disa are looking at the spoons. For simplicity, suppose you should be sure that the rational credence in top is either *high* or *low*:

$$(1) \quad S([P(\textit{Top}) = h] \vee [P(\textit{Top}) = l]) \quad (h > l)$$

Suppose that the rational credence is *high*:

$$(2) \quad P(\textit{Top}) = h$$

Since Disa is smart, learning that she’s *t*-confident of *Top* ($C_D(\textit{Top}) = t$) provides you with some reason to think that your evidence warrants having credence *t*:

$$(3) \quad P(P(\text{Top}) = t | C_D(\text{Top}) = t) > P(P(\text{Top}) = t)$$

Finally, since Disa's only route to the truth is through her evidence, once you learn what your shared evidence warrants, further learning Disa's actual opinion does not affect the probability of *Top*:

$$(4) \quad P(\text{Top} | P(\text{Top}) = t) = P(\text{Top} | [P(\text{Top}) = t] \wedge [C_D(\text{Top}) = s])$$

This is a paradigm peer-disagreement scenario. Alone, (1)–(4) do not require you to lower your confidence when you learn that Disa disagrees; but given Trust, they do:

Fact 8.1: (1)–(4) are consistent with $P(\text{Top} | C_D(\text{Top}) < h) \geq P(\text{Top})$. But given Trust, (1)–(4) imply $P(\text{Top} | C_D(\text{Top}) < h) < P(\text{Top})$.

Clearly this is only the beginnings of a theory of how to respond to disagreement. But it is a possibility proof: it shows that there *is* a principled way to take disagreement seriously. The next step is to find general rules delineating in what situations and to what extent Trust requires doing so. Trust gives answers—we just have to find them.

Upshot: Trust lays the foundations for a theory of disagreement.

8.2. To KK

In the wake of Williamson (2000) it has become popular to combine two views:

KNOWLEDGE FIRST

Knowledge plays the fundamental role in epistemology.

No KK

You can be in a position to know p without being in a position to know that you are ($Kp \wedge \neg KKp$).

Our theory puts pressure on this conjunction. For it is difficult to maintain Knowledge First without endorsing a close connection between knowledge and rational degrees of belief—after all, it is the latter that encode what you should think and determine what you should do. The natural way to forge this connection is through rational certainty: you should be sure of p iff you're in a position to know p ($Sp \leftrightarrow Kp$). But rationality must be something that we ought to treat as a guide. This, in turn, requires Trust and Value—both of which imply positive access:

Fact 8.2: *In any probability frame: if Value or Trust hold at a world, then $Sp \rightarrow SSp$ does as well.*

If knowledge is to be first, positive access in turn leads to KK: you are in a position to know p only if you're in a position to know that you are ($Kp \rightarrow KKp$).

Upshot: Trust forces a choice between Knowledge First and No KK.³⁰

³⁰ Knowledge First relies on a systematic and wide-ranging epistemological program. No KK rests, primarily, on subtle margin-for-error-like arguments (Williamson 2000; Bacon 2013) that can be resisted on a variety of grounds (Sharon and Spectre 2008; Greco 2014b; Stalnaker 2015; Das and Salow 2016). Given that we face a choice, I say that No KK should be the one to go.

9. A Modest Goal

I have a modest goal. Literally. My goal is an epistemology informed by modesty. Because you and I are modest: we are constantly wondering whether we are thinking and doing as we should. And we are not irrational for that. Given the sort of critters we are—and the sort of mistakes we make—we *should* have a healthy dose of self-doubt. My goal is an epistemology that makes room for critters like us—a modest epistemology.

In this paper, I've tried to secure its foundations. I argued that in order to take disagreement seriously, evidence must be a modest guide (§2); that the intuitive guiding principle—Reflection—rules out modesty entirely (§3); that its natural refinement—Trust—allows modesty while guaranteeing a correlation between evidence and truth (§§4–5); that denying Trust risks allowing you to know that your evidence is *anti*-correlated with truth (§6); that Trust characterizes the platitude that you should prefer to use free evidence (§7); and that the resulting theory *does* require you to take disagreement seriously (§8). In short, Trust demonstrates that there is a coherent, rational way to have a healthy dose of self-doubt. There are two directions to take it from here.

The first is toward extensions. The theoretical arguments of this paper pinpoint a particular structure of evidence. Structure in hand, we face both formal and philosophical questions. What is the *general* relation between Value, Trust, and that structure? And what sort of thing must evidence *be*, if it is to have that structure?

The second is toward applications. We now know how to model higher-order probabilities in a principled, coherent way. What does this tell us about...

- ... how to analyze disagreement and debunking (Christensen 2010a)?
- ... how to model “ambiguous” evidence (Joyce 2010)?
- ... how to rely on people who don't know their own opinions (Roush 2016)?
- ... how to defer to chances that don't know their own values (Briggs 2009)?

Foundations secured. Next step: developing modest answers to our new questions.³¹

Appendix A. Formal Details

In Appendix A I explain the formal details underlying the results stated in this paper. This section explains *probability frames* and how they encode higher-order information. §A.1 explains the special case of *prior frames*, how they can be used to characterize Trust, and how to interpret the tree-like structures that result. §A.2 explains how to generalize such frames to capture multiple bodies of evidence, and uses this generalization to characterize Value and Informed Trust. A **probability frame** $\langle W, \mathcal{P} \rangle$ consists of a finite set of worlds W and a function \mathcal{P} from worlds w to probability functions \mathcal{P}_w defined over the subsets of W (cf. Samet 1997; Dorst 2018a). \mathcal{P}_w represents the credences that you ought to have at w . Since rational uncertainty is modeled as uncertainty about which world you're in, and the rational credence function varies across worlds, you can have rational uncertainty about what the rational credence function is.

³¹ I can't thank everyone who has helped me with this long project. I received valuable feedback from Laura Callahan, Kenny Easwaran, Branden Fitelson, Brooke Husic, Harvey Lederman, Hanti Lin, Eric Pacuit, Miriam Schoenfield, Kieran Setiya, Jack Spencer, Steve Yablo, and many others. Special thanks to Bernhard Salow, Ginger Schultheis, Bob Stalnaker, and Roger White for sticking with me. And to Gillian Russell—for getting me started.

A proposition p is (any) set of worlds: $p \subseteq W$. p is true at w iff $w \in p$. Logical operations on propositions are captured with set-theoretic ones: $\neg p = W - p$; $p \wedge q = p \cap q$; etc. Probabilistic facts are captured in the obvious ways by using the probability functions associated with each world. Thus the proposition that *the rational credence in p is t* is the set of worlds w such that $\mathcal{P}_w(p) = t$: $[P(p) = t] =_{df} \{w | \mathcal{P}_w(p) = t\}$. Similarly for other probabilistic propositions. Since these higher-order claims are simply sets of worlds, they get assigned probabilities like any other proposition.

Example. Let $W = \{a, b\}$ with $\mathcal{P}_a(a) = 0.6$ (so $\mathcal{P}_a(b) = 0.4$) and $\mathcal{P}_b(a) = 0.3$ (so $\mathcal{P}_b(b) = 0.7$). We can diagram this with an arrow labeled t from x to y indicating that $\mathcal{P}_x(y) = t$, as in Figure 3.

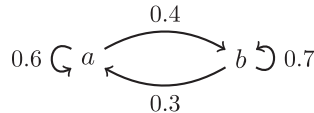


Fig. 3. A probability frame.

In this frame, $[P(a) = 0.6] = \{a\}$ (since $\mathcal{P}_a(a) = 0.6$ and $\mathcal{P}_b(a) = 0.3$), so in turn $[P(P(a) = .6) = 0.6]$ is the set of worlds w such that $\mathcal{P}_w(P(a) = .6) = 0.6$, i.e. the set of w such that $\mathcal{P}_w(\{a\}) = 0.6$, i.e. $\{a\}$ itself: $[P(P(a) = .6) = 0.6] = \{a\}$. Meanwhile, $[P(a) = 0.3] = \{b\}$ (at b , you should be 0.3 confident you're at a); $[P(P(a) = .6) = 0.3] = \{b\}$ (at b , you should be 0.3 confident that you should be 0.6 confident that you're at a); $[P(P(P(a) = .6) = .6) = 0.3] = \{b\}$ (at b , you should be 0.3 confident that you should be 0.6 confident that you should be 0.6 confident that you're at a); and so on. As can be seen, in such frames there can be rational uncertainty all the way up.

A.1. Trust the Details

To characterize Trust we will make use of a natural, tractable subclass of probability frames. Given a probability frame $\langle W, \mathcal{P} \rangle$, say that the evidence at w leaves open w' —and write wEw' —iff $\mathcal{P}_w(w') > 0$. The set of worlds that the evidence at w leaves open is $E_w =_{df} \{w' | wEw'\} = \{w' | \mathcal{P}_w(w') > 0\}$. A *prior frame* is a probability frame in which there is a unique prior π such that the probability function \mathcal{P}_w at each world w can be recovered by conditioning π on E_w . Precisely: a **prior frame** $\langle W, E, \pi \rangle$ is a probability frame $\langle W, \mathcal{P} \rangle$ in which there is a regular probability distribution π over W (i.e. $\forall w \in W: \pi(w) > 0$)—the prior—such that for all w : $\mathcal{P}_w = \pi(\cdot | E_w)$ (cf. Williamson 2000, 2014, 2018; Cresto 2012; Lasonen-Aarnio 2015; Salow 2017).

Formally, since the variation in probabilities across worlds is all traceable to the binary relation E , prior frames are based on a standard Kripke **frame** $\langle W, E \rangle$. This allows us to use the ordinary tools of modal logic to render them tractable.³² Philosophically, prior frames are natural given several different background pictures. These pictures all agree on two things: (1) rational opinion is separable into two components—(i) which standards of reasoning you should be using (π), and (ii) what you should take for granted (E_w); and (2) rational agents should know which standards of reasoning they should be using. Different approaches will fill out the components in different ways, but all will agree that what you should do is take that prior π and condition it on your evidence E_w .

³² Note: in any prior frame $\langle W, E, \pi \rangle$, E must be serial: $\forall x \exists y : xEy$.

And all can allow modesty—for even if you know what prior you should use, you can be uncertain what you should condition on. (Moreover, even if we reject such pictures, prior frames still capture an important case of modesty—so are a useful tool.)

We can now characterize Trust. First, some definitions. When xEy , I'll say “ x sees y .” A frame $\langle W, E \rangle$ is **transitive** iff: if x sees y and y sees z , then x sees z : $(xEy \wedge yEz) \Rightarrow xEz$. A frame is **shift-reflexive** iff any world seen by anything sees itself: $xEy \Rightarrow yEy$. A frame is **shift-nested** iff whenever two worlds are seen, either one sees everything the other does or they see nothing in common: $wEx, y \Rightarrow (E_x \subseteq E_y$ or $E_x \supseteq E_y$ or $E_x \cap E_y = \emptyset$). These conditions characterize Trust:

Theorem 5.7: (Trust Characterization) A prior frame $\langle W, E, \pi \rangle$ validates Trust iff $\langle W, E \rangle$ is transitive, shift-reflexive, and shift-nested.

Such frames are composed of structures like Figure 4. (Circles drawn around worlds that see exactly the same worlds; transitive arrows omitted.)

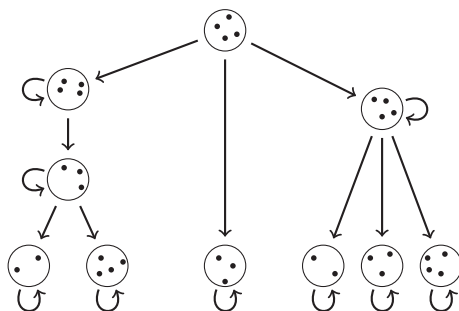


Fig. 4. A transitive, shift-reflexive, and shift-nested frame.

What we have is a *tree*.³³ The arrows can proceed down chains and splits, but they can never split and reconnect, as in Figure 5 below.

How do these constraints map onto the structure of evidence discussed in §5? Shift-reflexivity corresponds to surely-factivity: you should be sure that if you should be sure of p , then p is true; $S(Sp \rightarrow p)$. Transitivity corresponds to positive access: you should be sure of p only if you should be sure that you should be; $Sp \rightarrow SSp$.

Although shift-nesting does not correspond to an axiom that can be stated with the modal operator S , we can use other tools to get a grip on it. To see this, consider one interpretation of (a slight refinement of) our tree-structures. A frame is **reflexive** if every world sees itself: xEx . It is **nested** if for every pair of worlds, either one sees everything the other does or they see nothing in common: $(E_x \subseteq E_y$ or $E_x \supseteq E_y$ or $E_x \cap E_y = \emptyset$). We'll consider the set of transitive, reflexive, nested frames.

Think of the evidence as warranting not merely a set of conclusions, but also a *line of reasoning* to that set of conclusions. Lines of reasoning—like proofs—are path-dependent: “ p ; q ; therefore, $p \wedge q$ ” is a different line of reasoning than “ q ; p ; therefore, $p \wedge q$.” On this picture, you should leave open that a state might be the rational one only if it is an *extension* of the line of reasoning you should take. Since no two lines of reasoning

³³ Formally: if we take equivalence classes under E and force it irreflexive, the resulting structure of every E_w becomes a *forest* in graph-theoretic parlance.

can diverge and still wind up in the same place, this is why we get the “branching” structure. Precisely: say that $\langle W, E \rangle$ is **reasoning-generable** iff there is a function f from worlds w to sequences of propositions $f(w)$ such that: wEx iff $f(w)$ is an initial segment of $f(x)$. (Think of $f(w)$ as the sequence of conclusions that represent the line of reasoning you should take, given your evidence.) Then:

Theorem 5.8: $\langle W, E \rangle$ is reasoning-generable iff it is reflexive, transitive, and nested.

This gives some sense to our tree structures; but *why* would your evidence warrant concluding p earlier in your line of reasoning than concluding q ? Notice that if you have rationally concluded p but not q ($P(p) = 1$ but $P(q) < 1$), then conditional on only one of p and q being true, you should be certain it's p : $P(p|\neg(p \wedge q)) = 1$; you should think to yourself, “Supposing that at most one p or q is true, I'm sure it's p .” In other words, your epistemic access to p is *more robust* than your epistemic access to q . That may be why your evidence warrants concluding p earlier in your line of reasoning than concluding q .

We can reinforce this point with some independent resources. On most theories of conditionals, being conditionally certain of p given $\neg(p \wedge q)$ (i.e. $P(p|\neg(p \wedge q)) = 1$) implies that you should be certain of the indicative conditional, *If at most one of p or q is true, then it's p* . Notice that our certainties in such conditionals encode facts about how *epistemically robust* their components are. I'm certain that Oswald shot Kennedy. But I'm also certain that (1) *if Oswald didn't shoot Kennedy, then someone else did*. My certainty in (1) is no trivial consequence of the fact that I'm certain that the antecedent is false—for notice that I do *not* believe that *if Oswald didn't shoot Kennedy, then no one else did*. Rather, my certainty in (1) encodes the fact that my certainty that someone shot Kennedy is more robust than my certainty that Oswald did—if you were to remove my certainty of the latter, I'd still be sure of the former. Thinking of indicative conditionals as capturing epistemic dependence in this way makes two (widely-endorsed) theses extremely natural. Let ' $q \multimap p$ ' represent the indicative conditional *if q , then p* . Then:

- a) If $P(p|q) = 1$, then $P(q \multimap p) = 1$.
If conditional on q you should be sure of p , then you should be sure of *if q , p* .
- b) If $P(q \multimap p) = 1$, then $P(q \multimap \neg p) < 1$ (for $q \neq \emptyset$).
If you should be sure of *if q , p* , then you should not be sure of *if q , $\neg p$* .

(a) is an instance of the widely-attested Ramseyan thesis that the probability of an indicative conditional equals the corresponding conditional probability.³⁴ (b) is a consequence of the principle of “conditional non-contradiction” (not: *if q , then p* and *if q , then $\neg p$*), as well as the belief-revision principle that if you're certain of *if q , p* , then upon learning that q you should be certain of p (cf. Stalnaker 1984).

Say that a frame $\langle W, E \rangle$ is **conditionable** iff we can define a two-place propositional connective \multimap satisfying (a) and (b). Then in order to have a sensible connection between conditional-beliefs and beliefs in conditionals, we need a tree-structure:

³⁴ See Ramsey 1931; Stalnaker 1970; Adams 1975; van Fraassen 1976; Edgington 1995; Bennett 2003; Khoo 2013, 2016; Rothschild 2013; Bacon 2015. Fact 5.9 implies that no triviality results can be proved from (a); see footnote 35.

Fact 5.9: A reflexive, transitive $\langle W, E, \pi \rangle$ is conditionable iff it is nested.³⁵

To see why this is so, it will help to illustrate how Fact 5.9 furnishes a response to an objection to nesting (Das 2017; Williamson 2018). One type of higher-order uncertainty arises with a good/bad-case asymmetry: looking at a real painting, you can tell it's not a fake; but looking at a fake, you should be unsure whether it's the real thing (Lasonen-Aarnio 2015). Thus in the good case you can tell you're in the good case, but in the *bad* case you should be unsure which case you're in; omitting reflexive arrows, we have a structure like this: $b \rightarrow g$. Now suppose you are looking at *two* independent paintings. Shouldn't we expect a non-nested structure such as in Figure 5? (gb is where the first painting is real and the second is fake, etc; reflexive, transitive arrows omitted.)

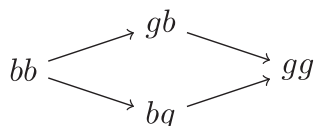
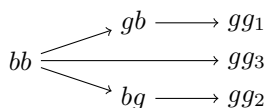


Fig. 5. A non-nested frame.

I object to this model, for it implies inconsistent comparisons of epistemic robustness. Notice that at gb , conditional on one of the paintings being fake, you should be certain that the first one is real: $\mathcal{P}_{gb}(gb|\neg gg) = 1$. Thus, applying (a), at gb you should be certain of the conditional, *if only one of them is real, it's the first one*: $\mathcal{P}_{gb}(\neg gg \rightarrow gb) = 1$. Since you are certain of this conditional and you leave open that you are at world gg , the conditional must be *true* at gg . And since at gg you should be certain you're at gg , it follows that at gg you should be certain of the conditional: $\mathcal{P}_{gg}(\neg gg \rightarrow gb) = 1$. Of course, at gg you should be certain that the antecedent of this conditional is false. But—as we saw with the Oswald example above—that does not trivialize your attitude; rather, your certainty in the conditional means that at gg your certainty that the first painting is real should be *more robust* than your certainty that the second one is. Here lies the problem: the cases are symmetric! Parallel reasoning starting from bg would lead to the opposite conclusion that in gg you should be certain that *if only one of them it real, it's the second one*; thus your certainty that the first one is real is *less* robust than your certainty that the second one is. Contradiction.

Fact 5.9 implies that to make this case consistent with principles (a) and (b), we must divide the gg -possibilities into ones that make true different facts about epistemic robustness. For instance, we may have one possibility (gg_1) where your evidential access to the first painting is more robust than your evidential access to the second, one (gg_2) where vice versa, and one (gg_3) where neither is more robust than the other:



³⁵ It follows from Fact 5.9 that no triviality results (à la Lewis 1976; Bradley 2000; Russell and Hawthorne 2016) can be proven from conditionability. Why? Because triviality results happen when you impose a connection between credences in conditionals and conditional credences for *all* probability functions (e.g. Bradley 2000). There is no threat if we only impose the connection for a *particular* probability function that is coordinated with the conditional (cf. Mandelkern and Khoo 2018)—as conditionability does.

I conclude that there is plenty of motivation—from Trust, conditionals, and (as we will see) the value of evidence—to defend nesting against the model given in Figure 5.

A.2. Value the Details

To formalize and then characterize Value, we must generalize our frames to allow for multiple bodies of evidence. A **dynamic probability frame** $\langle W, \mathcal{P}^1, \mathcal{P}^2 \rangle$ is a generalization of a probability frame. It associates a world w with *two* probability functions \mathcal{P}_w^1 and \mathcal{P}_w^2 . The intended interpretation is that \mathcal{P}_w^1 is the rational credence function given body 1 of evidence, and \mathcal{P}_w^2 is the rational credence function given body 2 of evidence—with 2 at least as informed as 1. (Thus we expect 1 to defer to itself and 2, but 2 to defer only to itself.) Propositions about probabilities are now indexed to their respective body of evidence; for $i = 1, 2$: $[P^i(p) = t] =_{df} \{w | \mathcal{P}_w^i(p) = t\}$. *Convention*: when using multiple variables to range over probability statements, I will use ***i* and *k*** with the constraint that $k \geq i$. Thus the statement $[P^i(P^k(p) = 0.7) \geq 0.4]$ expresses three claims: $[P^1(P^1(p) = 0.7) \geq 0.4]$, $[P^1(P^2(p) = 0.7) \geq 0.4]$, and $[P^2(P^2(p) = 0.7) \geq 0.4]$.

Given a dynamic probability frame $\langle W, \mathcal{P}^1, \mathcal{P}^2 \rangle$, write $wE^i w'$ iff $\mathcal{P}_w^i(w') > 0$. The set of worlds that evidence i leaves open at w is $E_w^i =_{df} \{w' | \mathcal{P}_w^i(w') > 0\}$. A **dynamic prior frame** $\langle W, E^1, E^2, \pi \rangle$ is a dynamic probability frame $\langle W, \mathcal{P}^1, \mathcal{P}^2 \rangle$ in which there is a regular probability distribution π over W such that for all w : $\mathcal{P}_w^i = \pi(\cdot | E_w^i)$.

Dynamic probability frames can interpret Value and Informed Trust. The latter is easy: it says $P_q^i(p | P_q^k(p) \geq t) \geq t$ (with $k \geq i$), and is valid iff all its instances are. How to formalize Value? Given a dynamic (probability or prior) frame F , we can enrich it with a **decision problem** $\langle O, U \rangle$: a set of options O and a real-valued utility function U .³⁶ $U_w(o)$ is the value of taking option o at world w . The *expected* value of an option o , according to evidence i , at world w , is an average of the various possible values of $U(o)$, with weights determined by how likely (according to \mathcal{P}_w^i) they are. Formally: $\mathbb{E}_w^i[U(o)] =_{df} \sum_{w'} \mathcal{P}_w^i(w') U_{w'}(o)$. What about the expected value, according to evidence i , of taking the option that is warranted by evidence k (whatever it is)? Since what k warrants believing (\mathcal{P}_w^k) varies across worlds, what k warrants *doing* does as well: if you're at w , then k warrants taking an option o that maximizes expected value by the lights of \mathcal{P}_w^k . So just as we have a function \mathcal{P}^k from worlds to probability functions that captures what you should *think* given evidence k , so too we'll have a function d that takes a body of evidence k and outputs a function d^k from worlds to options that captures what you should *do* given evidence k . Precisely: given a frame plus decision problem $\langle F, O, U \rangle$, d is (a variable over) any function from evidence k to a function d^k from worlds w to options $d_w^k \in O$ such that $d_w^k \in \arg \max_{o \in O} (\mathbb{E}_w^k[U(o)])$.³⁷ Thus the expected value relative to evidence i of doing what evidence k warrants is the expected value of taking option d^k (whatever it is): $\mathbb{E}_w^i[U(d^k)] =_{df} \sum_{w'} \mathcal{P}_w^i(w') U_{w'}(d_{w'}^k)$.

With these definitions we can formalize propositions about expected values in the obvious ways, e.g. $[\mathbb{E}^i[U(d^k)] \geq \mathbb{E}^i[U(o)]] =_{df} \{w | \mathbb{E}_w^i[U(d^k)] \geq \mathbb{E}_w^i[U(o)]\}$. Thus:

³⁶ Technical constraint: if W is the set of worlds in frame F , then $\langle O, U \rangle$ must be such that for any π over W , there is at least one option with maximal π -expected utility: $\max_{o \in O} (\mathbb{E}_\pi[U(o)])$ is non-empty. A sufficient (but not necessary) condition for this is for O to be finite.

³⁷ Since d captures a strategy of responding to evidence, we impose the constraint that it cannot vary across worlds where the relevant evidence warrants the exact same beliefs: if $\mathcal{P}_x^i = \mathcal{P}_y^i$, then $d_x^i = d_y^i$.

Value: $\mathbb{E}^i[U(d^k)] \geq \mathbb{E}^i[U(o)]$ ($o \in O, k \geq i$)

You should always expect taking the option warranted by informed evidence to be at least as good as you expect any other particular option to be.

Slogan: *evidence is valuable.*

As with Trust, Value has three types of instances depending on whether $i = 1 = k$, $i = 1$ and $k = 2$, or $i = 2 = k$. Value *holds* at a world w in a dynamic frame F iff for every decision problem $\langle O, U \rangle$: all instances of Value are true at w in $\langle F, O, U \rangle$. Value is *valid* iff it holds at all worlds.

What does it take to validate Value? In full generality, we have partial answers:

Fact 7.1: *In any dynamic probability frame: if Misguided Evidence or Effacing Evidence are true at a world, Value fails.*

Theorem 7.2: *In any dynamic probability frame: if Value holds at a world, Informed Trust does as well.*

Thus we know that, in full generality, Trust is necessary for Value. The converse remains an open question:

Conjecture 7.3: *In any dynamic probability frame: if Informed Trust holds at a world, then Value does as well.*

But within our tractable subclass—dynamic *prior* frames—the question is closed. First, some generalizations of our definitions. A dynamic frame $\langle W, E^1, E^2 \rangle$ is **transitive** iff $(xE^i y \wedge yE^i z) \Rightarrow xE^i z$. It is **shift-reflexive** iff $xE^i y \Rightarrow yE^k y$. It is **shift-nested** iff $wE^i x, y \Rightarrow (E_x^k \subseteq E_y^k \text{ or } E_x^k \supseteq E_y^k \text{ or } E_x^k \cap E_y^k = \emptyset)$. Finally, a dynamic frame is **shift-updating** iff if a world is seen by anything, it does not lose information between E^1 and E^2 : $xE^i y \Rightarrow E_y^2 \subseteq E_y^1$. Shift-updating corresponds to the surely-monotonicity axiom that you should be sure that if your evidence warrants being sure of p , a body of evidence at least as informed as your own will likewise warrant being sure of p : $S^i(S^i p \rightarrow S^k p)$.

These conditions characterize both Informed Trust and Value:

Theorem 7.4: (*Value and Trust*) *The following are equivalent:*

- (1) The dynamic prior frame $\langle W, E^1, E^2, \pi \rangle$ validates Informed Trust.
- (2) $\langle W, E^1, E^2 \rangle$ is transitive, shift-reflexive, shift-nested, and shift-updating.
- (3) The dynamic prior frame $\langle W, E^1, E^2, \pi \rangle$ validates Value.

Upshot: within dynamic prior frames, Trust characterizes the value of evidence.³⁸

³⁸ The implication from (2) to (3) is a generalization of [Geanakoplos 1989], Theorem 1—a result that was key to my discovery of the connection between Trust and Value.

Appendix B. Proofs

Fact 3.1: *If a probability frame validates Reflection, it validates $[P(\text{Immodest}) = 1]$.*

Proof: *Immodest* is true at x iff $\mathcal{P}_x(P = \mathcal{P}_x) = 1$. So it will suffice to take a probability frame that validates Reflection, suppose that $\mathcal{P}_w(P = \pi) > 0$, and then show that $\pi(P = \pi) = 1$. First suppose $\pi(P = \pi) = 0$; then $\mathcal{P}_w([P = \pi] \wedge [P(P = \pi) = 0]) > 0$, hence $\mathcal{P}_w(P = \pi | P(P = \pi) = 0) > 0$, contradicting Reflection. So $\pi(P = \pi) > 0$. We show that for an arbitrary world y such that $\pi(y) = t > 0$, it follows that $\pi(P(y) = t) = 1$. By Reflection, $\pi(y | P(y) = t) = t$, implying that $y \in [P(y) = t]$. But if $\pi(P(y) = t) < 1$, then $\pi(y | P(y) = t) = \frac{\pi(y \wedge [P(y) = t])}{\pi(P(y) = t)} = \frac{\pi(y)}{\pi(P(y) = t)} > \frac{\pi(y)}{1} = t$, i.e. $\pi(y | P(y) = t) > t$, contradicting Reflection. Since y was arbitrary: $\forall z \in W$: if $\pi(z) = t$, then $\pi(P(z) = t) = 1$. Since any probability function over W is fully determined by the probabilities it assigns to worlds, it follows that $\pi(P = \pi) = 1$. □

Remark 4.0: For any (finite) probability frame and any p, q, t , there is an $\epsilon > 0$ such that $[P_q(p) > t] = [P_q(p) \geq t + \epsilon]$. (Take the $w \in [P_q(p) > t]$ with minimal $\mathcal{P}_w(p|q)$ and set $\epsilon =_{df} (\mathcal{P}_w(p|q) - t)$.) So if $\langle W, \mathcal{P} \rangle$ validates Trust, it also validates $P_q(p|P_q(p) > t) > t$.

Fact 4.1: *In any probability frame: Trust holds at a world iff $P_q(p|P_q(p) \leq t) \leq t$ does.*

Proof: Note that $[P_q(p) \leq t] = [P_q(\neg p) \geq 1 - t]$. If Trust holds, then $P_q(\neg p | P_q(\neg p) \geq 1 - t) \geq 1 - t$, iff $P_q(\neg p | P_q(p) \leq t) \geq 1 - t$, iff $P_q(p | P_q(p) \leq t) \leq t$. Similarly for the converse. □

Fact 4.2: *There are probability frames that validate Trust in which Reflection fails at all worlds.*

Proof: The frame from Fact 5.1 validates Trust, yet it is easy to check that Reflection fails at all worlds. □

Fact 4.3: *In any probability frame, if Trust holds, and for a given $l, h \in [0, 1]$ it's true that $P(S(P(p) \in [l, h]) | P(p) \in [l, h]) = 1$, then $P(p | P(p) \in [l, h]) \in [l, h]$.*

Proof: Suppose Trust holds, and $P(S(P(p) \in [l, h]) | P(p) \in [l, h]) = 1$. Let $q =_{df} [P(p) \in [l, h]]$. Then $P_q(P = P_q) = 1$ (since conditional on q , you should be sure that the rational credence was already certain of q). Since Trust implies shift-reflexivity, the fact that $P_q(S(P(p) \in [l, h])) = 1$ implies that $P_q(P(p) \in [l, h]) = 1$; thus $P_q(P_q(p) \in [l, h]) = 1$. By Trust, $P_q(p) = P_q(p | P_q(p) \geq l) \geq l$, and $P_q(p) = P_q(p | P_q(p) \leq h) \leq h$, and so $P_q(p) = P(p | P(p) \in [l, h]) \in [l, h]$. □

Fact 5.1: *There are probability frames that validate both Trust and $[P(\text{modest}) = 1]$.*

Proof: Brute force method. Let $W = \{a, b\}$, $\mathcal{P}_a(a) = 0.6$ (so $\mathcal{P}_a(b) = 0.4$), and symmetrically $\mathcal{P}_b(b) = 0.6$ (so $\mathcal{P}_b(a) = 0.4$). $\text{modest} =_{df} \{w | 0 < \mathcal{P}_w(P = \mathcal{P}_w) < 1\}$. Since both \mathcal{P}_a and \mathcal{P}_b assign positive probability to $[P = \mathcal{P}_a]$ and $[P = \mathcal{P}_b]$, $\text{modest} = W$, hence $[P(\text{modest}) = 1] = W$. Moreover, this frame validates Trust. Since the frame is symmetric, it suffices to show that Trust holds at a . There are two nontrivial propositions: $\{a\}$ and $\{b\}$. If $q = \emptyset$, P_q is undefined at all worlds; if $q = \{a\}$ or $q = \{b\}$,

$P_q(a) = 1$ or $P_q(b) = 1$ (respectively), so Trust holds trivially. So we need only show that for $q = W$, Trust holds. For $t_1 \leq 0.4$, $[P(a) \geq t_1] = W$; hence $\mathcal{P}_a(a|P(a) \geq t_1) = \mathcal{P}_a(a) = .6 \geq t_1$. For $t_2 \in (0.4, 0.6]$, $[P(a) \geq t_2] = \{a\}$; hence $\mathcal{P}_a(a|P(a) \geq t_2) = \mathcal{P}_a(a|a) = 1 \geq t_2$. For $t_3 \in (0.6, 1]$, $[P(a) \geq t_3] = \emptyset$, so $\mathcal{P}_a(\cdot|P(a) \geq t_3)$ is undefined. Similarly: For $t_1 \leq 0.4$, $[P(b) \geq t_1] = W$; hence $\mathcal{P}_a(b|P(b) \geq t_1) = \mathcal{P}_a(b) = .4 \geq t_1$. For $t_2 \in (0.4, 0.6]$, $[P(b) \geq t_2] = \{b\}$; hence $\mathcal{P}_a(b|P(b) \geq t_2) = \mathcal{P}_a(b|b) = 1 \geq t_2$. For $t_3 \in (0.6, 1]$, $[P(b) \geq t_3] = \emptyset$, so $\mathcal{P}_a(\cdot|P(b) \geq t_3)$ is undefined. Thus Trust holds at a and, by symmetry, is valid. \square

Fact 5.2: For any $T = \{t_1, \dots, t_n\} \subset [0, 1]$, there are probability frames that validate Trust with a candidate π and proposition p such that $\pi(P(p) \in T) \approx 1$ and for all t_i : $\pi(P(p) = t_i) \approx \frac{1}{n}$. ($s \approx t$ means that $s \in [t - \epsilon, t + \epsilon]$, for arbitrarily small $\epsilon > 0$.)

Proof: Take any $T = \{t_1, \dots, t_n\} \subset [0, 1]$ and $\epsilon > 0$. By Theorem 5.7, it will suffice to construct a shift-reflexive, transitive, shift-nested prior frame $\langle W, E, \pi \rangle$ with the desired property. Let $W = \{w, a_1, b_1, \dots, a_n, b_n\}$. Define E so that $E_w = W$ while $E_{a_i} = E_{b_i} = \{a_i, b_i\}$. Clearly this frame is reflexive (every world sees itself), transitive (if xEy , $E_y \subseteq E_x$), and nested (if $E_x \not\subseteq E_y$ and $E_y \not\subseteq E_x$, then $E_x \cap E_y = \emptyset$), so any prior over it will validate Trust. In particular, let $\pi(w) = \epsilon$, and for each t_i , let $\pi(a_i) = \frac{t_i}{n}(1 - \epsilon)$ and $\pi(b_i) = \frac{1-t_i}{n}(1 - \epsilon)$. Then $\pi(W) = \pi(w) + \sum_{i=1}^n (\frac{t_i}{n} + \frac{1-t_i}{n})(1 - \epsilon) = \epsilon + \sum_{i=1}^n (\frac{1}{n})(1 - \epsilon) = \epsilon + (1 - \epsilon) = 1$, so π is probability function. Since $E_w = W$, $\mathcal{P}_w = \pi$; since $\mathcal{P}_w(w) > 0$, $\mathcal{P}_w = \pi$ will be our candidate. Let $p = \{a_1, \dots, a_n\}$. For each t_i , $\mathcal{P}_{a_i}(p) = \mathcal{P}_{b_i}(p) = \pi(a_i|\{a_i, b_i\}) = \frac{\frac{t_i}{n}(1-\epsilon)}{\frac{1}{n}(1-\epsilon)} = t_i$; hence $[P(p) = t_i]$ is true at $\{a_i, b_i\}$ (and so $[P(p) = t_j]$ for $t_j \neq t_i$ is false at $\{a_i, b_i\}$). In general for all $t_i \in T$: $[P(p) = t_i] = \{a_i, b_i\}$ unless $\mathcal{P}_w(p) = t_j$ for some $t_j \in T$, in which case $[P(p) = t_j] = \{a_j, b_j, w\}$. Since $\pi(w) = \epsilon$, it follows that $\pi(P(p) \in T) \geq \pi(\neg\{w\}) = 1 - \epsilon$. And for any $t_i \in T$: $\pi(P(p) = t_i) \geq \pi(\{a_i, b_i\}) = \frac{1}{n}(1 - \epsilon) = \frac{1}{n} - \frac{\epsilon}{n} \geq \frac{1}{n} - \epsilon$. Similarly, $\pi(P(p) = t_i) \leq \pi(\{a_i, b_i, w\}) = \frac{1}{n}(1 - \epsilon) + \epsilon \leq \frac{1}{n} - \frac{\epsilon}{n} + \epsilon \leq \frac{1}{n} + \epsilon$, as desired. \square

Fact 5.3: A probability frame validates Trust iff it validates both Reaction and Reliance.

Proof: (\Rightarrow): Suppose Trust is valid. Consider Reaction: if $P_q(P_q(p) \in [l, h]) = 1$ is true at a given world, then $P_q(p) = P_q(p|P_q(p) \geq l)$, which by Trust is $\geq l$. Similarly, $P_q(p) = P_q(p|P_q(p) \leq h)$, which by Trust and Fact 4.1 is $\leq h$. So $P_q(p) \in [l, h]$. Reaction is valid. Consider Reliance. Trust, Fact 4.1, and Remark 4.0 imply that $P_q(p|P_q(p) \geq t) \geq t > P_q(p|P_q(p) < t)$ whenever both are well-defined. By total probability, it follows that $P_q(p|P_q(p) \geq t) \geq P_q(p)$ whenever well-defined.

(\Leftarrow): Suppose Reaction and Reliance are valid. First note that Reaction requires that if $\mathcal{P}_w(x) > 0$, then $\mathcal{P}_x(x) > 0$ (shift-reflexivity). Two steps. First, if $\mathcal{P}_w(x) > 0$ and $\mathcal{P}_x(y) > 0$, then $\mathcal{P}_w(y) > 0$ (transitivity). For if not, then $\mathcal{P}_w(y|\{x, y\}) = 0$ even though $\mathcal{P}_w(P(y|\{x, y\}) > 0|\{x, y\}) = 1$. Now suppose $\mathcal{P}_w(x) > 0$ but $\mathcal{P}_x(x) = 0$. $\mathcal{P}_x(y) > 0$ for some y . If $\mathcal{P}_y(x) > 0$, by transitivity $\mathcal{P}_x(x) > 0$. If not, then $\mathcal{P}_w(P(x|\{x, y\}) = 0|\{x, y\}) = 1$, yet since $\mathcal{P}_w(x) > 0$ then $\mathcal{P}_w(x|\{x, y\}) > 0$, contra-

dicting Reaction. Now suppose $P_q(p|P_q(p) \geq t)$ is well-defined at an arbitrary x . Take arbitrary $w \in q \wedge [P_q(p) \geq t]$ such that $\mathcal{P}_x(w|q \wedge [P_q(p) \geq t]) > 0$. By shift-reflexivity, $\mathcal{P}_w(p|q \wedge [P_q(p) \geq t])$ is well-defined and by Reliance it is at least $P_w(p|q)$. Since $w \in [P_q(p) \geq t]$, it follows that $\mathcal{P}_w(p|q \wedge [P_q(p) \geq t]) \geq t$. Since w was arbitrary, $\mathcal{P}_x(P_q(p|q \wedge [P_q(p) \geq t]) \geq t | q \wedge [P_q(p) \geq t]) = 1$. Letting $r =_{df} q \wedge [P_q(p) \geq t]$, that is: $\mathcal{P}_x(P_r(p) \geq t|r) = 1$, so by Reaction, $\mathcal{P}_x(p|r) \geq t$, i.e. $P_q(p|P_q(p) \geq t) \geq t$ is true at x . Since x was arbitrary, Trust is valid. \square

Fact 5.4: *If Trust holds at a world in a probability frame, Misguided Evidence does not.*

Proof: Suppose $P(p \wedge [P(p) < t]) \geq t$. Since $P(P(p) < t) \leq 1$, it follows that $\frac{P(p \wedge [P(p) < t])}{P(P(p) < t)} \geq t$, i.e. $P(p|P(p) < t) \geq t$, violating Trust (Remark 4.0 and Fact 4.1). \square

Fact 5.5: *In any probability frame: if Trust holds then $[P(P(p) \geq t) \geq s] \rightarrow [P(p) \geq t \cdot s]$ does too. No stronger connection holds: for any $t, s \in [0,1]$: there are probability frames that validate Trust and make both $[P(P(p) \geq t) \geq s]$ and $[P(p) = t \cdot s]$ true at a world.*

Proof: Suppose Trust holds and $P(P(p) \geq t) \geq s$ is true. By Trust, $P(p|P(p) \geq t) \geq t$; by total probability, $P(p) = P(P(p) \geq t) \cdot P(p|P(p) \geq t) + P(P(p) < t) \cdot P(p|P(p) < t) \geq P(P(p) \geq t) \cdot P(p|P(p) \geq t) \geq s \cdot t$, so $P(p) \geq t \cdot s$.

Take any $t, s \in [0,1]$. Supposing $t, s \in (0,1)$, consider a prior frame $\langle W, E, \pi \rangle$ with $W = \{a, b, c\}$, $E_a = \{a, b, c\}$, $E_b = E_c = \{b, c\}$, and $\pi(a) = 1-s$, $\pi(b) = t \cdot s$, and $\pi(c) = (1-t)s$. This frame is reflexive, transitive, and nested, so by Theorem 5.7 it validates Trust. Letting $p = \{b\}$, $\mathcal{P}_w(p) = t \cdot s$, so $[P(p) = t \cdot s]$ is true at w . And $\mathcal{P}_b(p) = \mathcal{P}_c(p) = \frac{t \cdot s}{s} = t$, hence $\{b, c\} \subseteq [P(p) \geq t]$. Since $\mathcal{P}_w(\{b, c\}) = s$, it follows that $[P(P(p) \geq t) \geq s]$ is also true at w . Finally, if we allow t, s to be extremal, it is routine to modify the frame (dropping worlds as needed) to ensure that π is regular and hence that we have a prior frame. \square

Theorem 5.7: *(Trust Characterization). A prior frame $\langle W, E, \pi \rangle$ validates Trust iff $\langle W, E \rangle$ is transitive, shift-reflexive, and shift-nested.* \square

Proof: A prior frame $\langle W, E, \pi \rangle$ is a special case of a dynamic prior frame $\langle W, E^1, E^2, \pi \rangle$, setting $E^1 = E^2$. In this case, (1) Informed Trust is equivalent to Trust and (2) shift-updating is trivially satisfied; hence the result follows from Theorem 7.4.

Theorem 5.8: *$\langle W, E \rangle$ is reasoning-generable iff it is reflexive, transitive, and nested.*

Proof: (\Rightarrow): Suppose $\langle W, E \rangle$ is reasoning-generable. *Reflexivity:* Since $f(w)$ is an initial segment of $f(w)$, wEw . *Transitivity:* If $f(w)$ is an initial segment of $f(x)$ and $f(x)$ is an initial segment of $f(y)$, then $f(w)$ is an initial segment of $f(y)$. *Nested:* Given transitivity, a frame is nested iff if xEz and yEz , then xEy or yEx . And if $f(x)$ and $f(y)$ are both initial segments of $f(z)$, then one (or both) must be an initial segment of the other.

(\Leftarrow): For the converse, we need some definitions and a lemma.

Definition 5.8.a: E_y is an **expansion** of E_x iff $E_y \supset E_x$. And E_y is a **minimal expansion** of E_x iff E_y is an expansion of E_x , and for any expansion E_z of E_x , $E_z \supseteq E_y$.

Lemma 5.8.b: *If $\langle W, E \rangle$ is transitive and nested, then if E_x has an expansion, it has a minimal expansion.*

Proof: Suppose, for reductio, E_x has an expansion but has no minimal expansion. Thus:

$$(\forall E_y \supset E_x)(\exists E_z \supset E_x) \text{ such that } E_z \not\supseteq E_y \quad (*)$$

W (hence E_x) is finite, so suppose $|E_x| = n$. Since E_x has a positive but finite number of expansions (since $\wp(W)$ is finite), there must be a $k > n$ such that $(\forall E_y \supset E_x) |E_y| \geq k$ and $(\exists E_y \supset E_x) |E_y| = k$. Take some such E_y with $|E_y| = k$. By (*), there is an $E_z \supset E_x$ such that $E_z \not\supseteq E_y$. (Hence $E_z \neq E_y$.) If $E_z \subseteq E_y$, then since $E_z \neq E_y$, $E_z \subset E_y$, and hence E_z is an expansion of E_x with $|E_z| < |E_y| = k$. Contradiction. Thus $E_z \not\subseteq E_y$. Since also $E_z \not\supseteq E_y$, since E is nested it follows that $E_z \cap E_y = \emptyset$. Yet E_z and E_y are both expansions of E_x , so $E_z \cap E_y \supseteq E_x \neq \emptyset$. Contradiction. \square

Definition 5.8.c: Given Lemma 5.8.b, in transitive, nested frames we can define a function \mathcal{M} such that given E_x , $\mathcal{M}(E_x)$ is E_x 's minimal expansion (if it has an expansion). Define \mathcal{M}^n by induction: $\mathcal{M}^1(E_x) = \mathcal{M}(E_x)$, and $\mathcal{M}^{n+1}(E_x) = \mathcal{M}(\mathcal{M}^n(E_x))$. Let $\mathcal{E}(x) =_{df} \langle E_x, \mathcal{M}^1(E_x), \dots, \mathcal{M}^n(E_x) \rangle$ where $\mathcal{M}^n(E_x)$ has no expansion.

Lemma 5.8.d: *If $\langle W, E, \pi \rangle$ is transitive and nested, then every expansion E_z of E_x appears in $\mathcal{E}(x) = \langle E_0, E_1, \dots, E_n \rangle$.*

Proof: For reductio, suppose not: $E_z \supset E_x$ but E_z does not appear in $\mathcal{E}(x)$. Since $E_z \not\supseteq E_n$ (since E_n has no expansion) but $E_z \supset E_0 = E_x$, there must be an i such that $E_z \supset E_i$ but $E_z \not\supseteq E_{i+1}$. Since $E_z \neq E_{i+1}$, $\mathcal{M}(E_i) = E_y \neq E_z$. Since $E_y \cap E_z \supseteq E_x \neq \emptyset$, by nesting it follows that $E_y \supseteq E_z$ or $E_y \subseteq E_z$. Yet if $E_y \supseteq E_z$, then $E_y \supset E_z$ —and since $E_z \supset E_i$, it follows that E_y is *not* the minimal expansion of E_i after all: $\mathcal{M}(E_i) \neq E_y$. Contradiction. And if $E_y \subseteq E_z$, then $E_y \subseteq E_z$. Yet $E_y = E_{i+1}$, meaning $E_{i+1} \subseteq E_z$, contradicting what was established above. \square

We can now prove the right-to-left direction of Theorem 5.8. Suppose E is reflexive, transitive, and nested. Given any world x , define $f(x)$ to be the inverse of $\mathcal{E}(x)$, i.e. $f(x) = \langle \mathcal{M}^n(E_x), \dots, \mathcal{M}^1(E_x), E_x \rangle$. We want to show that for any x, y : yE_x iff $f(y)$ is an initial segment of $f(x)$. First suppose yE_x . By transitivity and reflexivity, $E_x \subseteq E_y$. If $y = x$, then of course $f(y)$ is an initial segment of $f(x)$. If $y \neq x$, then $E_{-y} \supset E_{-x}$, so E_y is an expansion of E_x . By Lemma 5.8.d, E_y appears in $\mathcal{E}(x) = \langle E_x, \dots, E_y, \mathcal{M}^1(E_y), \dots, \mathcal{M}^n(E_y) \rangle$, and hence $f(y) = \langle \mathcal{M}^n(E_y), \dots, E_y \rangle$ is an initial segment of $f(x)$. Conversely, suppose $f(y)$ is an initial segment of $f(x)$. Thus $f(x) = \langle \mathcal{M}^n(E_y), \dots, E_y, \dots, E_x \rangle$, and so $\mathcal{E}(x) = \langle E_x, \dots, E_y, \dots, \mathcal{M}^n(E_y) \rangle$, so E_y is an expansion of E_x : $E_x \subseteq E_y$, and so by reflexivity xEx and thus yEx . \square

Fact 5.9: *A reflexive, transitive $\langle W, E, \pi \rangle$ is conditionable iff it is nested.*

Proof: (\Rightarrow): Suppose we have a reflexive, transitive, non-nested frame; suppose, for reductio, that it is conditionable. Since nesting fails, we have x, y such that $E_x \not\subseteq E_y$ and $E_x \not\supseteq E_y$ and $E_x \cap E_y \neq \emptyset$. Let $C =_{df} E_x \cap E_y$. By reflexivity and transitivity we know

$E_x \supset C \subset E_y$. Thus $\mathcal{P}_x(E_x|\neg C) = 1$, so by conditionability $\mathcal{P}_x(\neg C \rightarrow E_x) = 1$, hence (since $C \subseteq E_x$) $C \subseteq [\neg C \rightarrow E_x]$. By transitivity, for any $c \in C$, $E_c \subseteq C$, so $\mathcal{P}_c(C) = 1$, so (α) : $\mathcal{P}_c(\neg C \rightarrow E_x) = 1$. Meanwhile, since $\mathcal{P}_y(E_y) = 1$, $\mathcal{P}_y(E_y|\neg C) = 1$, and so $\mathcal{P}_y(E_y \cap \neg C|\neg C) = 1$, and thus $\mathcal{P}_y(\neg E_x|\neg C) = 1$, implying $\mathcal{P}_y(\neg C \rightarrow \neg E_x) = 1$. Since $C \subseteq E_y$, $C \subseteq [\neg C \rightarrow \neg E_x]$, so $\mathcal{P}_c(\neg C \rightarrow \neg E_x) = 1$; contradicting (α) and the supposition that the frame is conditionable.

(\Leftarrow): Supposing we have a reflexive, transitive, nested frame, we can define $\mathcal{E}(x)$ as in Definition 5.8.c. Given w , let F_q^w be the first element of $\mathcal{E}(w)$ such that $F_q^w \cap q \neq \emptyset$, or $F_q^w = W$ if there is none. Now for any q, p , let $[q \rightarrow p] =_{df} \{w | F_q^w \cap q \subseteq p\}$. We show this obeys the properties of conditionability. **(a)**: Suppose $\mathcal{P}_w(p|q) = 1$. Then $E_w \cap q \neq \emptyset$ and $E_w \cap q \subseteq p$. Now consider any x such $x \in E_w$. By transitivity, $E_x \cap q \subseteq p$. Since E_w appears in $\mathcal{E}(x)$ and $E_w \cap q \neq \emptyset$, by Lemma 5.8.d, F_q^x appears no later than E_w in $\mathcal{E}(x)$. Thus E_w is an expansion of F_q^x , so $F_q^x \cap q \subseteq p$, and hence $x \in [q \rightarrow p]$. Since x was arbitrary, $\mathcal{P}_w(q \rightarrow p) = 1$. **(b)**: It'll suffice to show that for any x , if $x \in [q \rightarrow p]$ for $q \neq \emptyset$, then $x \notin [q \rightarrow \neg p]$. Supposing $x \in [q \rightarrow p]$, $F_q^x \cap q \subseteq p$. Since $F_q^x \cap q \neq \emptyset$, there is a $y \in p$: $y \in F_q^x \cap q$. So $F_q^x \cap q \not\subseteq \neg p$; hence $x \notin [q \rightarrow \neg p]$. \square

Fact 6.1: *If Trust holds at a world in a probability frame, Effacing Evidence does not.*

Proof: Suppose $S(p \leftrightarrow [P(p) < t])$ and $S(\neg p \leftrightarrow [P(p) > t])$ are true. If $P(p) = 0$, then $P(\neg p) = 1$ and hence $P(P(p) > t) = 1$; but by Trust and Remark 4.0, $P(p|P(p) > t) > t$, implying $P(p) > t \geq 0$ —contradiction. So $P(p) > 0$, and hence $P(P(p) < t) > 0$, so $P(p|P(p) < t)$ is well-defined. But since $P(p \leftrightarrow [P(p) < t]) = 1$, $P(p|P(p) < t) = 1 \not< t$, violating Trust (Fact 4.1). \square

Fact 7.1: *In any dynamic probability frame: if Misguided Evidence or Effacing Evidence are true at a world, Value fails.*

Proof: We prove the contrapositive. By Theorem 7.2, if Value holds at a world, then Informed Trust does as well. And the proofs of Facts 5.4 and 6.1 straightforwardly generalize to dynamic probability frames: if Informed Trust holds at a world, Misguided Evidence and Effacing Evidence do not.

Theorem 7.2: *In any dynamic probability frame: if Value holds at a world, Informed Trust does as well.*

Basic idea: Given an Informed Trust failure, we construct a conditional bet that you expect the informed evidence to warrant making a poor decision on.

Proof: Contraposing, suppose Informed Trust fails: for some p, q, t, i, k with $k \geq i$, $P_q^i(p|P_q^k(p) \geq t) < t$ is true at some world z . Thus $P_q^i(p|P_q^k(p) \geq t) = t - a$ for $a > 0$. Define our decision problem such that $O = \{n, b\}$; the *nope* option has 0 value at every world v , while the *b* is a conditional bet on p given q . For arbitrarily small $\epsilon > 0$:

$$U_v(n) = 0 \text{ for all } v \in W \quad U_v(b) = \begin{cases} 0 & \text{if } v \notin q \\ 1 - t + \epsilon & \text{if } v \in p \cap q \\ -t & \text{if } v \in \neg p \cap q \end{cases}$$

We first establish $(\alpha) : [P_q^k(p) \geq t] \subseteq [d^k = b]$ ($=_{df} \{w | d_w^k = b\}$).

Proof: Take arbitrary w such that $\mathcal{P}_w^k(p|q) \geq t$. Since it is well-defined, $\mathcal{P}_w(q) > 0$. We must show that $\mathbb{E}_w^k[U(n)] < \mathbb{E}_w^k[U(b)]$. $\mathbb{E}_w^k[U(n)] = 0$, of course. On the other hand,

$$\begin{aligned} \mathbb{E}_w^k[U(b)] &= \mathcal{P}_w^k(q) \cdot \mathbb{E}_w^k[U(b)|q] + \mathcal{P}_w^k(\neg q) \cdot \mathbb{E}_w^k[U(b)|\neg q] && \text{(Total Expectation)} \\ &= \mathcal{P}_w^k(q) \cdot \mathbb{E}_w^k[U(b)|q] && \text{(Since } \neg q \subseteq [U(b) = 0]) \end{aligned}$$

So it'll suffice to show that $\mathbb{E}_w^k[U(b)|q] > 0$.

$$\begin{aligned} \mathbb{E}_w^k[U(b)|q] &= \mathcal{P}_w^k(p|q) \cdot (1 - t + \epsilon) + \mathcal{P}_w^k(\neg p|q) \cdot (-t) \\ &\geq t \cdot (1 - t + \epsilon) + (1 - t) \cdot (-t) = t\epsilon > 0 && \text{(Since } \mathcal{P}_w^k(p|q) \geq t) \end{aligned}$$

Hence $\mathbb{E}_w^k[U(b)|q] > 0$, so $\mathbb{E}_w^k[U(b)] > 0 = \mathbb{E}_w^k[U(n)]$, and thus $d_w^k = b$. \square

Next we establish that $(\beta) : [P_q^k(p) < t] \subseteq [d^k = n]$.

Proof: Take any x such that $\mathcal{P}_x^k(p|q) < t$. Thus $\mathcal{P}_x^k(p|q) = t - d$ for some $d > 0$. (Since well-defined, $\mathcal{P}_x^k(q) > 0$.) Again, $\mathbb{E}_x^k[U(n)] = 0$; so we must show that $\mathbb{E}_x^k[U(b)] < 0$. Since $\mathbb{E}_x^k[U(b)|\neg q] = 0$ as before, we know $\mathbb{E}_x^k[U(b)] = \mathcal{P}_x^k(q) \cdot \mathbb{E}_x^k[U(b)|q]$. So it'll suffice to show that $\mathbb{E}_x^k[U(b)|q] < 0$.

$$\begin{aligned} \mathbb{E}_x^k[U(b)|q] &= \mathcal{P}_x^k(p|q) \cdot (1 - t + \epsilon) + \mathcal{P}_x^k(\neg p|q) \cdot (-t) \\ &= (t - d) \cdot (1 - t + \epsilon) + (1 - t + d) \cdot (-t) = \epsilon(t - d) - d \end{aligned}$$

As $\epsilon \rightarrow 0$ the left term vanishes and $\mathbb{E}_x^k[U(b)|q] < 0$. It follows that $\mathbb{E}_x^k[U(b)] < 0 = \mathbb{E}_x^k[U(n)]$, hence $d_x^k = n$. (Since $[P_q^k(p) < t]$ is finite, it follows that there is an ϵ small enough such that, for all $x \in [P_q^k(p) < t]$: $d_x^k = n$.) \square

We now turn to showing that at our original world z , $\mathbb{E}_z^i[U(d^k)] < 0 = \mathbb{E}_z^i[U(n)]$, and hence that Value fails at z . Since $\mathbb{E}_z^i[U(d^k)|\neg q] = 0$, we know

$$\mathbb{E}_z^i[U(d^k)] = \mathcal{P}_z^i(q) \cdot \mathbb{E}_z^i[U(d^k)|q]$$

with $\mathcal{P}_z^i(q) > 0$. So it'll suffice to show that $\mathbb{E}_z^i[U(d^k)|q] < 0$. Since $[P_q^k(p) \geq t]$ and $[P_q^k(p) < t]$ partition the q -worlds assigned positive probability by z (since Value requires shift-reflexivity³⁹, and hence that the conditional probability is well-defined), we know

$$\begin{aligned} \mathbb{E}_z^i[U(d^k)|q] &= \mathcal{P}_z^i(P_q^k(p) \geq t|q) \cdot \mathbb{E}_z^i[U(d^k)|q \wedge [P_q^k(p) \geq t]] \\ &\quad + \mathcal{P}_z^i(P_q^k(p) < t|q) \cdot \mathbb{E}_z^i[U(d^k)|q \wedge [P_q^k(p) < t]] \end{aligned}$$

(β) implies that $[P_q^k(p) < t] \subseteq [d^k = n]$, so $\mathbb{E}_z^i[U(d^k)|q \wedge [P_q^k(p) < t]] = 0$, hence the right summand drops out:

³⁹ If z sees w but w doesn't see itself, define a bet which pays off if $\neg\{w\}$ and has a huge cost $-N$ if w . Since \mathcal{P}_w will warrant taking the bet no matter how large N is, eventually $\mathbb{E}_z^i[U(d^k)] < 0$, and Value fails.

$$\mathbb{E}_z^i[U(d^k)|q] = \mathcal{P}_z^i(P_q^k(p) \geq t|q) \cdot \mathbb{E}_z^i[U(d^k)|q \wedge [P_q^k(p) \geq t]]$$

Thus it suffices to show that $\mathbb{E}_z^i[U(d^k)|q \wedge [P_q^k(p) \geq t]] < 0$. (α) implies that $[P_q^k(p) \geq t] \subseteq [d^k = b]$, so it'll in turn suffice to show that $\mathbb{E}_z^i[U(b)|q \wedge [P_q^k(p) \geq t]] < 0$. Since $[P_q^i(p)|P_q^k(p) \geq t] = t - a$ is true at z , $\mathcal{P}_z^i(p|q \wedge [P_q^k(p) \geq t]) = t - a$; hence

$$\begin{aligned} \mathbb{E}_z^i[U(b)|q \wedge [P_q^k(p) \geq t]] &= \mathcal{P}_z^i(p|q \wedge [P_q^k(p) \geq t]) \cdot (1 - t + \epsilon) + \mathcal{P}_z^i(\neg p|q \wedge [P_q^k(p) \geq t]) \cdot (-t) \\ &= (t - a) \cdot (1 - t + \epsilon) + (1 - t + a) \cdot (-t) = \epsilon(t - a) - a \end{aligned}$$

As $\epsilon \rightarrow 0$, the left term vanishes and $\mathbb{E}_z^i[U(b)|q \wedge [P_q^k(p) \geq t]] < 0$. It follows that $\mathbb{E}_z^i[U(d^k)|q \wedge [P_q^k(p) \geq t]] < 0$ and hence that $\mathbb{E}_z^i[U(d^k)|q] < 0$, and hence that $\mathbb{E}_z^i[U(d^k)] < 0 = \mathbb{E}_z^i[U(n)]$. Value fails at z . \square

Theorem 7.4: (Value and Trust) *The following are equivalent:*

- (1) The dynamic prior frame $\langle W, E^1, E^2, \pi \rangle$ validates Informed Trust.
- (2) $\langle W, E^1, E^2 \rangle$ is transitive, shift-reflexive, shift-nested, and shift-updating.
- (3) The dynamic prior frame $\langle W, E^1, E^2, \pi \rangle$ validates Value.

It is immediate from Theorem 7.2 that (3) implies (1); so we must establish two lemmas: Lemma 7.4.2 that (1) implies (2), and Lemma 7.4.3 that (2) implies (3). First, we extend our definitions of transitive, reflexive (etc.) to apply to sets of worlds:

Definition 7.4.1: Given a set $q \subseteq W$ in $\langle W, E^1, E^2 \rangle$ and arbitrary $x, y \in q$: q is **transitive** iff for any $z \in W$, $x \in E_z^k \Rightarrow E_x^k \subseteq E_z^k$; q is **reflexive** iff $x E^k x$; q is **updating** iff $E_x^2 \subseteq E_x^1$; q is **nested** iff $(E_x^k \subseteq E_y^k$ or $E_x^k \supseteq E_y^k$ or $E_x^k \cap E_y^k = \emptyset$). Note that $\langle W, E^1, E^2 \rangle$ is transitive, shift-reflexive, shift-nested, and shift-updating iff each E_w^i is transitive, reflexive, nested, and updating.

Now we prove that (1) implies (2) in Theorem 7.4.

Lemma 7.4.2: *The dynamic prior frame $\langle W, E^1, E^2, \pi \rangle$ validates Informed Trust only if $\langle W, E^1, E^2 \rangle$ is transitive, shift-reflexive, shift-nested, and shift-updating.*

Proof: We show the contrapositive, using Definition 7.4.1: supposing there is a w with E_w^i not transitive (etc.), we show that Informed Trust fails.

Transitivity: Suppose $\exists x \in E_w^i$ such that $x \in E_z^k$ but $E_x^k \not\subseteq E_z^k$. By regularity, $[P^k(E_z^k) < 1]$ is true at x . Since $\mathcal{P}_z^k(E_z^k) = 1$ and $z E^k x$, $\mathcal{P}_z^k(E_z^k | P^k(E_z^k) < 1) = 1$. Informed Trust fails at z (Fact 4.1).

Reflexivity: Suppose $\exists x \in E_w^i$ such that $\neg(x E^k x)$. Let $p = W - \{x\}$, so $[P^k(p) = 1] \wedge \neg p$ is true at x . Since $w E^i x$, $\mathcal{P}_w^i(\neg p | P^k(p) = 1) > 0$, so $\mathcal{P}_w^i(p | P^k(p) \geq 1) < 1$. Informed Trust fails.

Updating: We know E_w^i is reflexive. Suppose $\exists x \in E_w^i$ such that $E_x^2 \not\subseteq E_x^1$, so there is a y with $x E^2 y$ but $\neg(x E^1 y)$. By the latter, $\mathcal{P}_x^1(y) = 0$. By the former, $[P^2(y) > 0]$ is true at x . By reflexivity, $x E^1 x$, so $\mathcal{P}_x^1(P^2(y) > 0) > 0$. Combined, we have $\mathcal{P}_x^1(y | P^2(y) > 0) = 0$. Informed Trust fails.

Nesting: Suppose $\exists x, y, z \in E_w^i$ with $E_x^k \not\subseteq E_y^k$ and $E_x^k \not\supseteq E_y^k$, but $z \in E_x^k \cap E_y^k$. We know that E_w^i must be transitive, reflexive, and updating; we'll show that Informed Trust fails at w for $q = \{x, y, z\}$, $p = \{z\}$, and

$$t =_{df} \min_{v \in \{x, y, z\}} [\mathcal{P}_v^k(z|q)].$$

Since E^k is transitive, we know $xE^k y$, $yE^k x$, and $x, yE^k w$. By the definition of t , $[P_q^k(z) \geq t] \supseteq \{x, y, z\} = q \subseteq E_w^i$, so $(\alpha): q = q \cap [P_q^k(z) \geq t] \cap E_w^i$. Now, $x, y \notin E_z^k$ for otherwise $zE^k x$ or $zE^k y$, and so (by transitivity) $xE^k y$ or $yE^k x$ —contradiction. Thus:

$$\mathcal{P}_z^k(z|q) = \mathcal{P}_z^k(z|\{x, y, z\}) = 1. \tag{\beta}$$

Moreover, since $y \notin E_x^k$ and $x \notin E_y^k$:

$$\mathcal{P}_x^k(z|q) = \frac{\pi(z \cap E_x^k \cap q)}{\pi(E_x^k \cap q)} = \frac{\pi(z)}{\pi(\{x, z\})} \tag{\gamma}$$

$$\mathcal{P}_y^k(z|q) = \frac{\pi(z \cap E_y^k \cap q)}{\pi(E_y^k \cap q)} = \frac{\pi(z)}{\pi(\{y, z\})} \tag{\delta}$$

Combining (β) , (γ) , and (δ) , and the definition of t , we know

$$\begin{aligned} t &\geq \frac{\pi(z)}{\pi(\{x, z\})}, \frac{\pi(z)}{\pi(\{y, z\})} \\ &> \frac{\pi(z)}{\pi(\{x, y, z\})} && \text{(by regularity)} \\ &= \pi(z|q) = \pi(z|q \cap [P_q^k(z) \geq t] \cap E_w^i) \\ &= \mathcal{P}_w^i(p|q \cap [P_q^k(z) \geq t]) && \text{(by } (\alpha)) \end{aligned}$$

That is, $P_q^i(z|P_q^k(z) \geq t) < t$ at w : Informed Trust fails. □

The final step for Theorem 7.4 is showing that that (2) implies (3):

Lemma 7.4.3 (cf. Geanakoplos 1989): *If $\langle W, E^1, E^2 \rangle$ is transitive, shift-reflexive, shift-nested, and shift-updating, then $\langle W, E^1, E^2, \pi \rangle$ validates Value.*

Basic idea: Show that if the frame is transitive, shift-reflexive, shift-nested, and shift-updating, then we can partition E_w^i into smaller “branches,” with expectations from w an average of the expectations conditional on each branch. Then an induction on the size of E_w^i suffices to carry Value from the “leaves” up through the tree.

Definition 7.4.3.a (k-closed): $q \subseteq W$ is k -closed iff for any $x \in q$, $E_x^k \subseteq q$. Note that if E_w^i is transitive and updating, E_w^i is k -closed. (Updating implies that for any $x \in E_w^i$, then $E_x^2 \subseteq E_x^1$, so $E_x^k \subseteq E_x^i$, and by transitivity $E_x^i \subseteq E_w^i$.)

Definition 7.4.3.b (k-classes): Given a k -closed $q \subseteq W$, let its set \mathcal{N}^k of k -classes partition q into worlds that see the same worlds under E^k : $\mathcal{N}^k =_{df} \{N \subseteq q | \forall x, y \in N : E_x^k = E_y^k\}$.

The k -class of a world x is denoted $N_x^k = \{y \in q \mid E_y^k = E_x^k\}$. We let A^k denote the k -class whose members see All of q under E^k : $A^k =_{df} \{x \in q \mid E_x^k = q\}$. (A^k may be empty.)

Definition 7.4.3.c (k -class accessibility): If q is transitive, reflexive, and k -closed, and $N, M \in \mathcal{N}^k$, then $(\exists n \in N, m \in M : nE^k m)$ iff $(\forall n \in N, m \in M : nE^k m)$. (Why? Suppose $nE^k m$. $\forall m' \in M$: (reflexivity) $m'E^k m'$, so (same k -class) $mE^k m'$, so (transitivity) $nE^k m'$; so (same k -class) $\forall n' \in N : n'E^k m'$.) Thus within E_w^i we can treat E^k as a relation between k -classes: for $N, M \in \mathcal{N}^k : NE^k M$ iff $\exists n \in N, m \in M : nE^k m$, iff $\forall n \in N, m \in M : nE^k m$. Similarly for the neighborhood of a class: $E_N^k =_{df} \{x \in q \mid \exists y \in N : E_y^k = E_x^k\} = \{x \in q \mid \forall y \in N : E_y^k = E_x^k\}$. Note: by reflexivity and transitivity: $NE^k M$ iff $E_M^k \subseteq E_N^k$; and $E_M^k \subseteq E_N^k$ iff $NE^k M$ and $N \neq M$.

Definition 7.4.3.d (Maximal k -classes): Given a transitive, reflexive, and k -closed q , the maximal k -classes of q are those that see strictly less than q under E^k but are not seen by any other k -classes that do so: $\{M \in \mathcal{N}^k \mid E_M^k \subset q \text{ and } \neg \exists K \in \mathcal{N}^k : E_M^k \subset E_K^k \subset q\}$.

Fact 7.4.3.e: If q is transitive, reflexive, nested, and k -closed, and M_1, \dots, M_n are its maximal k -classes, then it is partitioned by $\{A^k, E_{M_1}^k, \dots, E_{M_n}^k\}$.

Proof: **Exhaustivity:** Take arbitrary $x \in q$. By reflexivity, $x \in N_x^k$. If $E_{N_x^k}^k \not\subset q$, then $E_{N_x^k}^k = q$, so $x \in A^k$, hence covered by $\{A^k, E_{M_1}^k, \dots, E_{M_n}^k\}$. So suppose $E_{N_x^k}^k \subset q$; we show that $x \in E_{M_j}^k$ for some maximal M_j . By reflexivity $x \in E_x^k$, so $N_x^k E^k N_x^k$. Therefore there must be a class M_j that's maximal and $M_j E^k N_x^k$. For suppose not: there is no $K \in \mathcal{N}^k$ such that $E_K^k \subset q$, $KE^k N_x^k$, and (by definition of maximal) $\neg \exists K' \in \mathcal{N}^k : E_K^k \subset E_{K'}^k \subset q$, i.e.

$$\forall K \in \mathcal{N}^k : \text{if } E_K^k \subset q \text{ and } KE^k N_x^k \text{ then } \exists K' \in \mathcal{N}^k : E_K^k \subset E_{K'}^k \subset q. \quad (\alpha)$$

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

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

Definition 7.4.3.f: Given a dynamic prior frame $\langle W, E^1, E^2, \pi \rangle$ and any random variable X , let $\mathbb{E}_q[X]$ ($= \mathbb{E}[X|q]$) be the expectation of X relative to π conditional on q : $\mathbb{E}_q[X] =_{df} \sum_t \pi(X = t|q) \cdot t$. And let $\pi_q =_{df} \pi(\cdot|q)$.

Fact 7.4.3.g: If q is k -closed, transitive, reflexive, and nested, then any k -closed $r \subseteq q$ is also transitive, reflexive, and nested.

Proof: Let $x, y \in r$. *Transitive:* Suppose $\exists z \in W$ such that $x \in E_z^k$. Since $x \in q$ and q is transitive, if $x E^k y$ then $z E^k y$. *Reflexive:* Since $x \in q$ and q is reflexive, $x E^k x$. *Nested:* since $x, y \in q$, either $E_x^k \subseteq E_y^k$ or $E_x^k \supseteq E_y^k$ or $E_x^k \cap E_y^k = \emptyset$. \square

Lemma 7.4.3.h: If q is k -closed, transitive, reflexive, and nested, then for any decision problem $\langle O, U \rangle$ and any d and $o \in O$: $\mathbb{E}_q[U(d^k)] \geq \mathbb{E}_q[U(o)]$.

Proof: We proceed by induction on the size of q . *Base case:* If $|q| = 1$, then $q = \{x\}$. Since q is k -closed and reflexive, $E_x^k = \{x\}$, so $\pi(\cdot|q) = \mathcal{P}_x^k$; so for any random variable X , $\mathbb{E}_q[X] = \mathbb{E}_x^k[X]$. Moreover $\pi(d^k = d_x^k|q) = 1$, so $\mathbb{E}_q[U(d^k)] = \mathbb{E}_q[U(d_x^k)] = \mathbb{E}_x^k[U(d_x^k)] = \max_{o \in O} (\mathbb{E}_x^k[U(o)]) = \max_{o \in O} (\mathbb{E}_q[U(o)])$; hence $\mathbb{E}_q[U(d^k)] \geq \mathbb{E}_q[U(o)]$.

Induction case: Suppose $|q| = n$ and for all $r \subseteq W$ with $|r| < |q|$, the hypothesis holds. By Fact 7.4.3.e, if M_1, \dots, M_l are q 's maximal k -classes, then q can be partitioned by $\{A^k, E_{M_1}^k, \dots, E_{M_l}^k\}$. We thus can break down $\mathbb{E}_q[U(o)]$ as follows:

$$\mathbb{E}_q[U(o)] = \pi_q(A^k)\mathbb{E}_q[U(o)|A^k] + \sum_j \pi_q(E_{M_j}^k)\mathbb{E}_q[U(o)|E_{M_j}^k] \tag{\alpha}$$

Supposing $A^k = \emptyset$, the first summand drops out:

$$\mathbb{E}_q[U(o)] = \sum_j \pi_q(E_{M_j}^k)\mathbb{E}_q[U(o)|E_{M_j}^k] \tag{\beta}$$

Since each $E_{M_j}^k \subset q$ and q is transitive and k -closed, $E_{M_j}^k$ is k -closed. By Fact 7.4.3.g, $E_{M_j}^k$ is also transitive, reflexive, and nested. Since it is smaller than q , the inductive hypothesis holds and $\mathbb{E}[U(d^k)|E_{M_j}^k] \geq \mathbb{E}[U(o)|E_{M_j}^k]$. And since $q \cap E_{M_j}^k = E_{M_j}^k$, for any random variable X , $\mathbb{E}[X|E_{M_j}^k] = \mathbb{E}_q[X|E_{M_j}^k]$. Plugging these facts into (β) yields:

$$\leq \sum_j \pi_q(E_{M_j}^k)\mathbb{E}_q[U(d^k)|E_{M_j}^k] = \mathbb{E}_q[U(d^k)]$$

That is, $\mathbb{E}_q[U(o)] \leq \mathbb{E}_q[U(d^k)]$, as desired.

Next suppose $A^k \neq \emptyset$, so we have some $w \in A^k$ such that $E_w^k = q$. Then $\pi_q = \mathcal{P}_w^k$, $\mathbb{E}_q[X] = \mathbb{E}_w^k[X]$; and so by the definition of d_w^k we have:

$$\begin{aligned} \mathbb{E}_q[U(o)] &= \mathbb{E}_w^k[U(o)] \leq \mathbb{E}_w^k[U(d_w^k)] \\ &= \mathcal{P}_w^k(A^k)\mathbb{E}_w^k[U(d_w^k)|A^k] + \sum_j \mathcal{P}_w^k(E_{M_j}^k)\mathbb{E}_w^k[U(d_w^k)|E_{M_j}^k] \end{aligned} \tag{\gamma}$$

Since for any $x \in A^k$, $\mathcal{P}_x^k = \mathcal{P}_w^k$, we know $d_x^k = d_w^k$. Hence $\mathbb{E}_w^k[U(d_w^k)|A^k] = \mathbb{E}_w^k[U(d^k)|A^k]$. And by parallel reasoning to above, for each $E_{M_j}^k$, $\mathbb{E}_w^k[U(d_w^k)|E_{M_j}^k] \leq \mathbb{E}_w^k[U(d^k)|E_{M_j}^k]$. These facts imply a comparison with (γ) :

$$\leq \mathcal{P}_w^k(A^k)\mathbb{E}_w^k[U(d^k)|A^k] + \sum_j \mathcal{P}_w^k(E_{M_j}^k)\mathbb{E}_w^k[U(d^k)|E_{M_j}^k] = \mathbb{E}_w^k[U(d^k)]$$

And since $\mathbb{E}_w^k[U(d^k)] = \mathbb{E}_q[U(d^k)]$, we have the desired result. \square

We are finally in a position to complete the proof of Theorem 7.4 by establishing Lemma 7.4.3.

Lemma 7.4.3 (cf. Geanakoplos 1989): *If $\langle W, E^1, E^2 \rangle$ is transitive, shift-reflexive, shift-nested, and shift-updating, then $\langle W, E^1, E^2, \pi \rangle$ validates Value.*

Proof: Suppose $\langle W, E^1, E^2 \rangle$ is transitive, shift-reflexive, shift-nested, and shift-updating, and consider an arbitrary prior frame $\langle W, E^1, E^2, \pi \rangle$ built on it. Consider an arbitrary world w . By Definitions 7.4.1 and 7.4.3.a, E_w^i is k -closed, transitive, reflexive, and nested. Thus Lemma 7.4.3.h applies: for any decision problem $\langle O, U \rangle$ and any d and $o \in O$: $\mathbb{E}[U(d^k)|E_w^i] \geq \mathbb{E}[U(o)|E_w^i]$; that is, $\mathbb{E}_w^i[U(d^k)] \geq \mathbb{E}_w^i[U(o)]$: Value holds. \square

Recall our disagreement setup:

- (1) $S([P(\text{Top})=h] \vee [P(\text{Top})=l])$ (h>l)
- (2) $P(\text{Top}) = h$
- (3) $P(P(\text{Top}) = t | C_D(\text{Top}) = t) > P(P(\text{Top}) = t)$
- (4) $P(\text{Top} | [P(\text{Top}) = t] \wedge [C_D(\text{Top}) = s]) = P(\text{Top} | P(\text{Top}) = t)$.

Fact 8.1: *(1)–(4) are consistent with $P(\text{Top} | C_D(\text{Top}) < h) \geq P(\text{Top})$. But given Trust, (1)–(4) imply $P(\text{Top} | C_D(\text{Top}) < h) < P(\text{Top})$.*

Proof: Our model of (1)–(4) will be based on the Sycophants frame. $W = \{s, p\}$. $\mathcal{P}_s(s) = 0.1$ and $\mathcal{P}_s(p) = 0.9$. Similarly, $\mathcal{P}_p(p) = 0.1$ and $\mathcal{P}_p(s) = 0.9$. Set $\text{Top} = p$ and $h = 0.9$. At world s , $S([P(p) = 0.9] \vee [P(p) = 0.1])$ and $[P(p) = 0.9]$ are true, so (1) and (2) are satisfied. Enrich this model with a function \mathcal{D} from worlds to Disa’s credences; and set $\mathcal{D}(w) = \mathcal{P}_w$ for each w . Then $S([C_D(q) = t] \leftrightarrow [P(q) = t])$ is valid. Thus $P(P(p) = t | C_D(p) = t) = 1 > .9 \geq P(P(p) = t)$, so (3) is satisfied. Finally, since $S([P(p) = t] \leftrightarrow [C_D(p) = t])$ is valid, $P(p | [P(p) = t] \wedge [C_D(p) = s]) = P(p | P(p) = t)$ if well-defined, so (4) is satisfied. Nevertheless, $P(p | C_D(p) < .9) = P(p | P(p) < .9) = 1 \geq .9 = P(p)$, establishing the first result.

Now suppose Trust holds and (1)–(4) are true. Let $q = [C_D(\text{Top}) < h]$. Then $P(\text{Top} | C_D(\text{Top}) < h) = P_q(\text{Top})$ can be broken down:

$$= P_q(P(\text{Top}) = l)P_q(\text{Top} | P(\text{Top}) = l) + P_q(P(\text{Top}) = h)P_q(\text{Top} | P(\text{Top}) = h)] \tag{by (1)}$$

$$= P_q(P(\text{Top}) = l)P(\text{Top} | P(\text{Top}) = l) + P_q(P(\text{Top}) = h)P(\text{Top} | P(\text{Top}) = h) \tag{by (4)}$$

From (3) it follows that $P(P(\text{Top}) = l | C_D(\text{Top}) < h) = P_q(P(\text{Top}) = l) > P(P(\text{Top}) = l)$ and $P_q(P(\text{Top}) = h) > P(P(\text{Top}) = h)$. Since

$$P(\text{Top}) = P(P(\text{Top}) = l)P(\text{Top} | P(p) = l) + P(P(\text{Top}) = h)P(\text{Top} | P(p) = h)$$

$P_q(\text{Top})$ is weighted more towards $P(\text{Top} | P(p) = l)$ than $P(\text{Top})$ is. Since by (1), $P(\text{Top} | P(\text{Top}) = l) = P(\text{Top} | P(\text{Top}) \leq l) \leq l$ (by Trust), and $l < h \leq P(\text{Top} | P(\text{Top}) \geq h)$

$= P(\text{Top}|\text{P}(\text{Top}) = h)$, it follows that $P_q(\text{Top})$ is more weighted towards the lower value than $P(\text{Top})$ is, so $P_q(\text{Top}) = P(\text{Top}|C_D(\text{Top}) < h) < P(\text{Top})$. \square

Fact 8.2: *In any probability frame: if Value or Trust hold at a world, then $Sp \rightarrow SS_p$ does as well.*

Proof: By Theorem 7.2, it suffices to show the result for Trust. For reductio, suppose Sp and $\neg SS_p$ are true, so $[P(p) = 1]$ yet $P(p|P) = 1 < 1$. By the latter, $P(p|P) < 1$ is well-defined; by the former, $P(p|P) < 1 = 1$, violating Trust. \square

Appendix C. Glossary

This glossary collects brief definitions of the technical terms used in the statements of principles and theorems, in alphabetical order.

- **Candidate:** π is a *candidate* in frame $\langle W, \mathcal{P} \rangle$ iff at some world you should think π might be the rational credence function: $\exists w \in W : \mathcal{P}_w(P = \pi) > 0$.
- **Conditionable:** A frame $\langle W, E \rangle$ is *conditionable* iff we can define a two-place propositional connective \rightarrow such that (a) $P(p|q) = 1$ implies $P(q \rightarrow p) = 1$, and (b) for $q \neq \emptyset$, $P(q \rightarrow p) = 1$ implies $P(q \rightarrow \neg p) < 1$.
- **d :** Given a frame plus decision problem $\langle F, O, U \rangle$, d is (a variable over) any function from evidence k to a functions d^k from worlds w to options $d_w^k \in O$ such that (1) $d_w^k \in \arg \max_{o \in O} (\mathbb{E}_w^k[U(o)])$ and (2) if $\mathcal{P}_x^i = \mathcal{P}_y^k$, then $d_x^i = d_y^k$.
- **Decision Problem $\langle O, U \rangle$:** A set of options O and a real-valued utility function U ; $U_w(o)$ is the utility of $o \in O$ at w .⁴⁰
- **Dynamic Prior Frame $\langle W, E1, E2, \pi \rangle$:** A dynamic probability frame $\langle W, \mathcal{P}^1, \mathcal{P}^2 \rangle$ in which there is a regular probability distribution π over W —the prior—such that for all w : $\mathcal{P}_w^i = \pi(\cdot | E_w^i)$.
- **Dynamic Probability Frame $\langle W, \mathcal{P}^1, \mathcal{P}^2 \rangle$:** A probability frame with an extra function \mathcal{P}^2 ; each possibility w is associated with a credence function \mathcal{P}_w^1 that's rational given evidence 1 and \mathcal{P}_w^2 that's rational given evidence 2.
- **Effacing Evidence:** $\exists p, t: S(\leftrightarrow [P(p) < t])$ and $S(\neg p \leftrightarrow [P(p) > t])$.
- $\mathbb{E}_w^i[U(o)] =_{df} \sum_{w'} \mathcal{P}_w^i(w') U_{w'}(o)$
- $\mathbb{E}_w^i[U(d^k)] =_{df} \sum_{w'} \mathcal{P}_w^i(w') U_{w'}(d_{w'}^k)$
- **$wE^i w'$, E_w^i :** Given a dynamic probability frame $\langle W, \mathcal{P}^1, \mathcal{P}^2 \rangle$, $wE^i w'$ iff $\mathcal{P}_w^i(w') > 0$ and $E_w^i =_{df} \{w' | \mathcal{P}_w^i(w') > 0\}$. In probability frames the superscripts are omitted.
- **Frame:** A structure $\langle W, E \rangle$ underlying a prior frame $\langle W, E, \pi \rangle$. W is finite; E is serial: $\forall x \exists y : xEy$.
- **Holds:** A principle *holds* at world w iff all of its instances (well-defined instantiations of free variables) are true at w .
- **i and k :** *Convention:* when using multiple variables over bodies of evidence, i and k obey the constraint that $k \geq i$.
- **Informed:** A body of evidence is *informed* (with respect to yours) iff it contains all your evidence, and maybe more.

⁴⁰ For $\langle O, U \rangle$ to enrich a frame with set of worlds W , $\langle O, U \rangle$ must be such that for any π over W there is at least one option with maximal π -expected utility: $\max_{o \in O} (\mathbb{E}_\pi[U(o)])$ is non-empty.

- **Informed Trust:** $P_q^i(p|P_q^k(p) \geq t) \geq t$ ($k \geq i$)
- **Judgment:** You *judge* that p iff you have a credence in p above the given (contextually specified) threshold t .
- **Misguided Evidence:** $\exists p, t : P(p \wedge [P(p) < t]) \geq t$
- **Nested:** A frame $\langle W, E \rangle$ is *nested* iff for any x, y : either $E_x \subseteq E_y$ or $E_x \supseteq E_y$ or $E_x \cap E_y = \emptyset$.
- **Opinion:** You have a given *opinion* about p iff you have a credence in p in the given (contextually specified) range $[l, h]$.
- π : A probability function whose values (unlike P) are known. $\pi_q =_{df} \pi(\cdot|q)$.
- \mathcal{P}_w : The probability function which is rational at world w ; it's values are known.
- $[P^i(p) = t]$: Given a dynamic probability frame, $[P^i(p) = t] =_{df} \{w | \mathcal{P}_w^i(p) = t\}$. In probability frames the superscript is omitted.
- $P_q(p)$ ($= P(p|q)$): The rational credence—whatever it is—in p conditional on q .
- **Prior Frame $\langle W, E, \pi \rangle$:** A probability frame $\langle W, \mathcal{P} \rangle$ in which there is a regular probability distribution π over W (i.e. $\forall w \in W: \pi(w) > 0$) such that for all w : $\mathcal{P}_w = \pi(\cdot|E_w)$.
- **Probability Frame $\langle W, \mathcal{P} \rangle$:** A structure consisting of a finite set of epistemic possibilities W and a function \mathcal{P} that associates each possibility w with the rational credences \mathcal{P}_w that you should have at w .
- **Reaction:** If $P_q(P_q(p) \in [l, h]) = 1$, then $P_q(p) \in [l, h]$
- **Reasoning-generable:** A frame $\langle W, E \rangle$ is *reasoning-generable* iff there is a function f from worlds w to sequences of propositions $f(w)$ such that: wEx iff $f(w)$ is an initial segment of $f(x)$.
- **Reflection:** $P(p|P(p) \in [l, h]) \in [l, h]$
- **Reliance:** $P_q(p|P_q(p) \geq t) \geq P_q(p)$
- **Reflexive:** A frame $\langle W, E \rangle$ is *reflexive* iff for all x : xEx .
- **Shift-Nested:** A dynamic frame $\langle W, E^1, E^2 \rangle$ is *shift-nested* iff $wE^i x, y \Rightarrow (E_x^k \subseteq E_y^k$ or $E_x^k \supseteq E_y^k$ or $E_x^k \cap E_y^k = \emptyset$). (For a frame $\langle W, E \rangle$, omit the superscripts.)
- **Shift-Reflexive:** A dynamic frame $\langle W, E^1, E^2 \rangle$ is *shift-reflexive* iff $xE^i y \Rightarrow yE^k y$. (For a frame $\langle W, E \rangle$, omit the superscripts.)
- **Shift-Updating:** A dynamic frame $\langle W, E^1, E^2 \rangle$ is *shift-updating* iff $xE^i y \Rightarrow E_y^2 \subseteq E_y^1$.
- **Sp:** Sp iff you should be Sure of p , iff $P(p) = 1$.
- **Transitive:** A dynamic frame $\langle W, E^1, E^2 \rangle$ is *transitive* iff $(xE^i y \wedge yE^i z) \Rightarrow xE^i z$. (For a frame $\langle W, E \rangle$, omit the superscripts.)
- **Trust:** $P_q(p|P_q(p) \geq t) \geq t$
- **Validates:** A frame *validates* a principle iff the principle is true at all worlds for all well-defined instantiations of its free variables.
- **Value:** $\mathbb{E}^i[U(a^k)] \geq \mathbb{E}^i[U(o)]$ ($o \in O, k \geq i$)

References

Adams, Ernest, 1975. *The Logic of Conditionals*, volume 86 of *Synthese Library*. Springer Netherlands.

- Ahmed, Arif and Salow, Bernhard, 2018. 'Don't Look Now'. *British Journal for the Philosophy of Science*
- Bacon, Andrew, 2013. 'Stalnaker on the KK principle'. Manuscript.
- Bacon, Andrew, 2015. 'Stalnaker's Thesis in Context'. *The Review of Symbolic Logic*, 8 (1):131–163.
- Bennett, Jonathan, 2003. *A Philosophical Guide to Conditionals*. Clarendon Press.
- Bradley, Richard, 2000. 'A preservation condition for conditionals'. *Analysis*, 60 (267):219–222.
- Briggs, Ray, 2009. 'The Anatomy of the Big Bad Bug'. *Noûs*, 43(3):428–449.
- Christensen, David, 2007. 'Epistemology of Disagreement: The Good News'. *Philosophical Review*, 116(2):187–217.
- , 1991. 'Clever Bookies and Coherent Beliefs'. *Philosophical Review*, 100(2):229–247.
- , 2010a. 'Higher-Order Evidence'. *Philosophy and Phenomenological Research*, 81 (1):185–215.
- , 2010b. 'Rational Reflection'. *Philosophical Perspectives*, 24:121–140.
- , 2016. 'Disagreement, Drugs, etc.: From Accuracy to Akrasia'. *Episteme*.
- Coates, Allen, 2012. 'Rational Epistemic Akrasia'. *American Philosophical Quarterly*, 49 (2):113–124.
- Cresto, Eleonora, 2012. 'A Defense of Temperate Epistemic Transparency'. *Journal of Philosophical Logic*, 41(6):923–955.
- Das, Nilanjan, 2017. 'Externalism and the Value of Information'. Manuscript.
- and Salow, Bernhard, 2016. 'Transparency and the KK Principle'. *Noûs*.
- Dorst, Kevin, 2017. 'Lockeans Maximize Expected Accuracy'. *Mind*, To Appear.
- , 2018a. 'Higher-Order Uncertainty'. In Mattias Skipper Rasmussen and Asbjørn Steglich-Petersen, eds., *Higher-Order Evidence: New Essays*, To appear. Oxford University Press.
- , 2018b. *Modest Epistemology*. Ph.D. thesis, Massachusetts Institute of Technology.
- Easwaran, Kenny, 2013. 'Expected Accuracy Supports Conditionalization – and Conglomerability and Reflection'. *Philosophy of Science*, 80(1):119–142.
- , Fenton-Glynn, Luke, Hitchcock, Christopher, and Velasco, Joel D, 2016. 'Updating on the Credences of Others: Disagreement, Agreement, and Synergy'. *Philosophers' Imprint*, 16(11).
- Edgington, Dorothy, 1995. 'On Conditionals'. *Mind*, 104:235–329.
- Elga, Adam, 2007. 'Reflection and Disagreement'. *Noûs*, 41(3):478–502.
- , 2010. 'Subjective probabilities should be sharp'. *Philosophers' Imprint*, 10(5):1–11.
- , 2013. 'The puzzle of the unmarked clock and the new rational reflection principle'. *Philosophical Studies*, 164(1):127–139.
- Feldman, Richard, 2005. 'Respecting the Evidence'. *Philosophical Perspectives*, 19 (1):95–119.
- , 2007. 'Reasonable religious disagreements'. In Louise Antony, ed., *Philosophers Without Gods: Meditations on Atheism and the Secular*, 194–214. Oxford University Press.
- Foley, Richard, 1992. 'The Epistemology of Belief and the Epistemology of Degrees of Belief'. *American Philosophical Quarterly*, 29(2):111–124.
- , 2009. 'Beliefs, Degrees of Belief, and the Lockean Thesis'. In Franz Huber and Christoph Schmidt-Petri, eds., *Degrees of Belief*, 37–47. Springer.

- Gaifman, Haim, 1988. 'A Theory of Higher Order Probabilities'. In Brian Skyrms and William L Harper, eds., *Causation, Chance, and Credence*, volume 1, 191–219. Kluwer.
- Geanakoplos, John, 1989. 'Game Theory Without Partitions, and Applications to Speculation and Consensus'. *Research in Economics*, Cowles Foundation (914).
- Gibbons, John, 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, Daniel, 2014a. 'A puzzle about epistemic akrasia'. *Philosophical Studies*, 161:201–219.
- , 2014b. 'Could KK be OK?' *Journal of Philosophy*, 111(4):169–197.
- Hall, Ned, 1994. 'Correcting the Guide to Objective Chance'. *Mind*, 103(412):505–517.
- Hazlett, Allan, 2012. 'Higher-order epistemic attitudes and intellectual humility'. *Episteme*, 9(3):205–223.
- Horowitz, Sophie, 2013. 'Immoderately rational'. *Philosophical Studies*, 167(1):41–56.
- , 2014. 'Epistemic Akrasia'. *Noûs*, 48(4):718–744.
- , 2018. 'Predictably Misleading Evidence'. In Mattias Skipper and Asbjørn Steglich-Petersen, eds., *Higher-Order Evidence: New Essays*, To appear. Oxford University Press.
- Huemer, Michael, 2011. 'The Puzzle of Metacoherence'. *Philosophy and Phenomenological Research*, 82(1):1–21.
- Huttegger, Simon M, 2014. 'Learning experiences and the value of knowledge'. *Philosophical Studies*, 171(2):279–288.
- Joyce, James M, 1998. 'A Nonpragmatic Vindication of Probabilism'. *Philosophy of Science*, 65(4):575–603.
- , 2009. 'Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief'. In Franz Huber and Christoph Schmidt-Petrii, eds., *Degrees of Belief*, 263–297. Springer.
- , 2010. 'A Defense of Imprecise Credences in Inference and Decision Making'. *Philosophical Perspectives*, 24(1):281–323.
- Kelly, Thomas, 2010. 'Peer disagreement and higher order evidence'. In Alvin I Goldman and Dennis Whitcomb, eds., *Social Epistemology: Essential Readings*, 183–217. Oxford University Press.
- Khoo, Justin, 2013. 'Conditionals, Indeterminacy, and Triviality'. *Philosophical Perspectives*, 27(1):260–287.
- , 2016. 'Probabilities of Conditionals in Context'. *Linguistics and Philosophy*, 39(1):1–43.
- Lasonen-Aarnio, Maria, 2010. 'Unreasonable Knowledge'. *Philosophical Perspectives*, 24(1):1–21.
- , 2013. 'Disagreement and evidential attenuation'. *Noûs*, 47(4):767–794.
- , 2014. 'Higher-order evidence and the limits of defeat'. *Philosophy and Phenomenological Research*, 8(2):314–345.
- Lasonen-Aarnio, Maria, 2015. 'New Rational Reflection and Internalism about Rationality'. In Tamar Szabó Gendler and John Hawthorne, eds., *Oxford Studies in Epistemology*, volume 5, 145–171. Oxford University Press.
- Leitgeb, Hannes, 2013. 'Reducing Belief Simpliciter to Degrees of Belief'. *Annals of Pure and Applied Logic*, 164:1338–1389.

- Levinstein, Ben, 2017. 'Permissive Rationality and Sensitivity'. *Philosophy and Phenomenological Research*, 94(2):343–370.
- Lewis, David, 1976. 'Probabilities of Conditionals and Conditional Probabilities'. *The Philosophical Review*, 85(3):297–315.
- , 1980. 'A Subjectivist's Guide to Objective Chance'. In Richard C Jeffrey, ed., *Studies in Inductive Logic and Probability*, volume 2. University of California Press.
- Littlejohn, Clayton, 2015. 'Stop Making Sense? On a Puzzle about Rationality'. *Philosophy and Phenomenological Research*, To Appear.
- Mandelkern, Matthew and Khoo, Justin, 2018. 'Against Preservation'. *Analysis*, To appear.
- Myrvold, Wayne C, 2012. 'Epistemic Values and the Value of Learning'. *Synthese*, 187(2):547–568.
- Oddie, Graham, 1997. 'Conditionalization, Cogency, and Cognitive Value'. *The British Journal for the Philosophy of Science*, 48(4):533–541.
- Paul, LA, 2014. *Transformative Experience*. Oxford University Press.
- Pettigrew, Richard, 2013. 'Epistemic Utility and Norms for Credences'. *Philosophy Compass*, 8(10):897–908.
- , 2016. *Accuracy and the Laws of Credence*. Oxford University Press.
- and Titelbaum, Michael G, 2014. 'Deference Done Right'. *Philosopher's Imprint*, 14(35):1–19.
- Ramsey, FP, 2010. 'Truth and probability'. In Antony Eagle, ed., *Philosophy of Probability: Contemporary Readings*. Routledge.
- Ramsey, Frank, 1931. *The Foundations of Mathematics and Other Logical Essays*. Kegan Paul, Trench, and Trubner & Co.
- Rothschild, Daniel, 2013. 'Do Indicative Conditionals Express Propositions?' *Noûs*, 47(1):49–68.
- Roush, Sherrilyn, 2009. 'Second Guessing: A Self-Help Manual'. *Episteme*, 251–268.
- , 2016. 'Knowledge of Our Own Beliefs'. *Philosophy and Phenomenological Research*, 93(3).
- Russell, Jeffrey Sanford and Hawthorne, John, 2016. 'General Dynamic Triviality Theorems'. *Philosophical Review*, 125(3):307–339.
- Salow, Bernhard, 2017. 'The Externalist's Guide to Fishing for Compliments'. *Mind*, To appear.
- Samet, Dov, 1997. 'On the Triviality of High-Order Probabilistic Beliefs'. <https://ideas.repec.org/p/wpa/wuwpga/9705001.html>.
- Schoenfield, Miriam, 2014. 'Permission to Believe: Why Permissivism is True and What it Tells Us About Irrelevant Influences On Belief'. *Noûs*, 48(2):193–218.
- , 2015a. 'A Dilemma for Calibrationism'. *Philosophy and Phenomenological Research*, 91(2):425–455.
- , 2015b. 'Bridging Rationality and Accuracy'. *Journal of Philosophy*, 112(12):633–657.
- , 2016. 'An Accuracy Based Approach to Higher Order Evidence'. *Philosophy and Phenomenological Research*, To Appear.
- Sharon, Assaf and Spectre, Levi, 2008. 'Mr. Magoo's Mistake'. *Philosophical Studies*, 139(2):289–306.
- Skyrms, Brian, 1966. *Choice and Chance: An Introduction to Inductive Logic*, volume 18. Dickenson Pub. Co.

- , 1980. 'Higher Order Degrees of Belief'. In D H Mellor, ed., *Prospects for Pragmatism*, 109–137. Cambridge University Press.
- , 1990. 'The Value of Knowledge'. *Minnesota Studies in the Philosophy of Science*, 14:245–266.
- Sliwa, Paulina and Horowitz, Sophi, 2015. 'Respecting *all* the evidence'. *Philosophical Studies*, 172(11):2835–2858.
- Smithies, Declan, 2012. 'Moore's paradox and the accessibility of justification'. *Philosophy and Phenomenological Research*, 85(2):273–300.
- , 2015. 'Ideal Rationality and Logical Omniscience'. *Synthese*, 192(9):2769–2793.
- Stalnaker, Robert, 1970. 'Probability and Conditionals'. *Philosophy of Science*, 37(1):64–80.
- , 1984. *Inquiry*. Cambridge University Press.
- , 2006. 'On the Logics of Knowledge and Belief'. *Philosophical Studies*, 128 (1):169–199.
- , 2015. 'Luminosity and the KK Thesis'. In Sanford Goldberg, ed., *Externalism, Self-Knowledge, and Skepticism*, 1–19. Cambridge University Press.
- Sturgeon, Scott, 2008. 'Reason and the Grain of Belief'. *Noûs*, 42(1):139–165.
- Tal, Eyal, 2018. 'Self-Intimation, Infallibility, and Higher-Order Evidence'. *Erkenntnis*, To appear.
- Titelbaum, Michael, 2015. 'Rationality's Fixed Point (or: In Defense of Right Reason)'. In Tamar Szabó Gendler and John Hawthorne, eds., *Oxford Studies in Epistemology*, volume 5, 253–292. Oxford University Press.
- and Kopec, Matthew, 2017. *Plausible Permissivism*. Oxford University Press.
- van Fraassen, Bas, 1976. 'Probabilities of Conditionals'. In William Harper and Clifford Hooker, eds., *Foundations of probability theory, statistical inference, and statistical theories of science.*, 261–308. Reidel.
- , 1984. 'Belief and the Will'. *The Journal of Philosophy*, 81(5):235–256.
- Vavova, Katia, 2014. 'Confidence, Evidence, and Disagreement'. *Erkenntnis*, 79:173–183.
- , 2016. 'Irrelevant Influences'. *Philosophy and Phenomenological Research*, To appear.
- Wedgwood, Ralph, 2012. 'Justified Inference'. *Synthese*, 189:273–295.
- White, Roger, 2005. 'Epistemic Permissiveness'. *Philosophical Perspectives*, 19(1):445–459.
- , 2009a. 'Evidential Symmetry and mushy credence'. *Oxford Studies in Epistemology*, 5: 161–186.
- , 2009b. 'On Treating Oneself and Others as Thermometers'. *Episteme*, 6(3):233–250.
- Williamson Timothy, 2000. *Knowledge and its Limits*. Oxford University Press.
- , 2014. 'Very Improbable Knowing'. *Erkenntnis*, 79(5):971–999.
- , 2018. 'Evidence of Evidence in Epistemic Logic'. In Mattias Skipper and Asbjørn Steglich-Petersen, eds., *Higher-Order Evidence: New Essays*, To appear. Oxford University Press.
- Worsnip, Alex, 2015. 'The Conflict of Evidence and Coherence'. *Philosophy and Phenomenological Research*, To Appear.