When Epistemic Models Misfire:
Lessons for Everyday Rationality[1]

Scott Sturgeon

1. Background.

Three natural assumptions are often built into formal models of rationality. The assumptions interact to generate faulty prescription for extreme credence. This suggests that one of two things is happening: either one (or more) of the assumptions is faulty somehow, or, failing that, one (or more) of them is not well formalized in our models. To see which it is we must do three things: examine how assumptions about rationality are built into our formal frameworks, examine how assumptions interact to generate faulty prescription for credence, and diagnose the resulting situation. Once we have done these things important lessons can be learned about everyday rationality.

This is the type of rationality manifested by ordinary people in everyday life. It stands in sharp contrast, of course, with the usual target of formal work in the area—ideal rationality—something never manifested by ordinary people in everyday life. The main goal of this paper is to reflect on situations in which formal models misfire in a certain way and use that reflection better to understand everyday rationality.

To begin, the first assumption about rationality is the view is that agents possess sources of evidence which plump for pieces of information as correct. On this view, evidence consists in claims or thoughts or other alethic items—items in the business of being accurate or inaccurate—and the key idea is that total evidence *in situ* consists in a collection of such items put forward by an agent's evidential sources. This is our first assumption about rationality

(1) Agents have sources of evidential information.

Whatever these sources turn out to be—states of perceptual experience, states of reflection on concepts or meaning, testimony by others, chunky facts in the nearby world—the thought is that evidential sources plump for alethic items as correct, and this impinges on propositional attitudes subject to rationality.[2]

[1] Many thanks to Dave Chalmers, Cian Dorr, Ant Eagle, Salvatore Florio, Nick Jones, Mark Kaplan, Matt Parrott, Richard Pettigrew, Susanna Siegel, Maja Spener, Timothy Williamson, and Big-Al Wilson for help with the paper.

[2] Since alethic items are objects of knowledge, our first assumption about rationality is consistent with the view that evidence is fixed by what is known. It is also consistent with non-factive approaches to evidence on which evidence comes from something like testimony of the senses (or testimony of others). What all these approaches to evidence have in common—with common-sense, of course, and crucial to our story later—is the idea that evidence is content-laden and *upstream* from attitudes subject to epistemic rationality. See note 18 for further discussion. For a knowledge-based conception of evidence see Timothy Williamson's *Knowledge and Its Limits* (OUP 2000), and for further discussion of that conception see contributions to Peter Greenough and Duncan Pritchard's *Williamson on Knowledge* (OUP 2009). For relevant sensory-based conceptions of evidence see contributions to Berit Brogard's *Does Perception Have Content?* (OUP 2014), Susanna Schellenberg's *The Unity of Perception: Content, Consciousness, Evidence* (OUP 2018), or Susanna Siegel's *The Rationality of Perception* (OUP 2016). For discussion of testimonial evidence see Elizabeth Fricker's "The Epistemology of Testimony" *Proceedings of the Aristotelian Society* 1987.

Our second assumption concerns how bits of information relate to one another. To a rough first approximation, it is the view that bits of information form into stories which support, cut against, or remain neutral about pieces of information. For any claim *C* and piece of information *I*, the key thought is that *C* has an *evidential standing* on *I*. Put with our first assumption the second can be rendered as

(2) Claims enjoy evidential standing on the total evidence.

This sort of standing might be strong—if evidence entails a claim, for instance—or it might be weak—if evidence entails a claim's negation—or it might be in-between in strength—if evidence is less relevant or committal about whether a given claim is true. In this way our second assumption is that claims enjoy levels of support on the total evidence.

Our third assumption is that strength of credence should match strength of evidential standing on the total evidence. When a claim is strongly or weakly or neutrally supported by evidence, an agent's credence for it should likewise be strong or weak or middling in character. We have here a matching hypothesis

(3) Strength of credence should match evidential support on the total evidence.

In a nutshell, then, our three starting assumptions are that agents possess sources of evidential information, claims enjoy evidential standing on the total evidence, and strength of credence should match that evidential standing. A natural formalization of these assumptions leads directly to bother. It sheds useful light on the relation between everyday and ideal rationality.

2. Formalization.

On the assumption that evidence comes in pieces of information, we can see those pieces as coming in collections. This means we can use the language of set theory to formalize when a piece of information is part of the total evidence. When *I* is information got from an evidential source, and **TE** is an agent's total evidence, our first assumption about rationality can be formalized as

$$(1)^{+} \quad I \in \mathbf{TE}.$$

This codifies the idea that information got from an evidential source is part of the total evidence.

If bits of information have evidential standing in light of the total evidence, however, it is natural to see that standing as coming in *degrees*. This idea moves past the more blunt view that bits of information have evidential standing on the total evidence, for we're assuming here that said standing comes in grades or shades. But this further thought is routine in formal epistemology and fits well with common-sense.

Once degrees of evidential support are on the table, though, it is natural to see them as shaped by logical relations.[3] Whenever logic alone ensures that exactly one of some claims is true, therefore, it is natural to think that levels of support for each claim on the evidence sums to maximal support. To get a feel for how this works, consider a pair of examples:

[3] For discussion of this idea see David Christensen's *Putting Logic in Its Place* (OUP 2004).

- No matter how total evidence turns out, support for P plus support for $\neg P$ adds-up to maximal support. There is always maximal support for $(P \vee \neg P)$, after all, since $(P \vee \neg P)$ is a tautology with exclusive disjuncts. This means its support intuitively adds-up from exclusive and exhaustive ways the disjunctive claim can be true: either because P is true, or because $\neg P$ is true.

- No matter how total evidence turns out, support for $(A \& C)$ plus support for $(A \& \neg C)$ plus support for $\neg A$ adds-up to maximal support. There is always maximal support for $[(A \& C) \vee (A \& \neg C) \vee \neg A]$, after all, since the claim is a tautology with exclusive disjuncts. This means its support intuitively adds-up from exclusive and exhaustive ways the disjunctive claim can be true: either because A is true with C, because A is true without C, or because A is not true.

And so on.

Once this sort of adding-up of support is in place, our second assumption about rationality is easy to formalize. If support works as the bullet points suggest, after all, evidence works like a probability function. Our second assumption about rationality may then be formalized as

$$(2)^{+} \qquad \underline{\text{ES}}(h, \mathbf{TE}) \;=\; \underline{\text{Ev-Prob}}(h/\mathbf{TE}).$$

No matter what the hypothesis h or the total evidence **TE**, the thought is that evidential support for h on **TE** is identical to the evidential probability of h given **TE**. In a phrase: evidential support is probability conditional on total evidence.

Our final assumption about rationality is the view that credence should match evidential support on total evidence. No matter what the hypothesis h or the total evidence **TE**: if we let $\underline{\text{Ev-Prob}}(\text{-})$ be a probabilistic measure of evidential support, and we let "≅" stand for when there rationally should be identity-of-value, then, our final assumption is

$$(3)^{+} \qquad \text{cr}(h) \;\cong\; \underline{\text{Ev-Prob}}(h/\mathbf{TE}).$$

This is the view that no matter what the hypothesis h or total evidence **TE**, credence for h should be identical to the evidential probability of h on the total evidence **TE**.

Pulling it all together, then, when formal models of rationality are put together it is often assumed that evidence comes in bits of information, that evidential support works like a probability function, and that strength of credence should match probability on the total evidence. $(1)^{+}$, $(2)^{+}$, and $(3)^{+}$ formalize these ideas. Together they lead to difficulty.

3. Three Cases.

Models endorsing $(1)^{+}$, $(2)^{+}$, and $(3)^{+}$ seem to misfire in a range of well-known cases. Each time they do so the models prescribe extreme credence for something which seems not to merit it. To see how this works, let ☺ be a claim which stands no epistemic risk of being false—a super-simple truth of logic, perhaps, or the claim that everything is self-identical, or whatever your favourite no-risk claim happens to be. Then difficulties crop up in at least three kinds of case.

The logic case.

Suppose *Log* is a complicated truth of logic, something it would take a normal person days to understand, much less prove. Our formalism entails that an agent is rational only if she is certain that *Log* is true. After all, its evidential support function, together with its conversion of evidential support to credal strength, jointly ensure maximal credence for *Log*:

$$[(2)^{+} \,\&\, (3)^{+}] \;\Rightarrow\; \mathrm{cr}(Log) \cong 1.$$

This means our formalism treats *Log* on a par with ☺: assigning both maximal credence. Since ☺ enjoys no epistemic risk of error, while *Log* isn't obvious at all, a natural thought is that this result is unacceptable. The palpable asymmetry between ☺ and *Log* suggests that a model of rationality is acceptable only if it marks the asymmetry somehow. Our formalism looks to fail this test, however, since it characterizes a person's take on a claim with credence, and it prescribes equal credence for ☺ and for *Log*.[4]

The coin case.

Suppose a fair coin is tossed an infinite number of times. Let *Hog* be the claim that it lands heads-on-the-ground every time. Our formalism entails that a person is rational only if she lends no credence to *Hog*. After all, its evidential support function, together with its take on the conversion of evidential strength to credal strength, jointly prescribe minimal credence for *Hog*:

$$[(2)^{+} \,\&\, (3)^{+}] \;\Rightarrow\; \mathrm{cr}(Hog) \cong 0.$$

This can easily seem wrong, since our formalism treats *Hog* on a par with the negation of our risk-free claim ☺, assigning both minimal credence. A natural thought is that a model of rationality is acceptable only if it marks the fact that rational agents should take a different stand on *Hog* and on ¬☺. Rational agents recognize that ¬☺ has no epistemic chance of being true, after all, yet they see plainly that *Hog* might well be true, as it describes one of the many ways that the coin might play out. Our formalism fails to mark the palpable asymmetry between ¬☺ and *Hog*, treating them as if they're on an epistemic par.[5]

The dog case.

Suppose a source of evidence endorses the claim that there is a dog before you. Perhaps it looks as if there is a dog before you, or it sounds as if there is a dog before you, or someone (you trust) says that there is a dog before you. Let *Dog* be the claim so endorsed; then our formalism entails that you should be certain of *Dog*. The claim is a piece of your total evidence, after all, so our formalism's evidential support function, together with its view about the conversion of evidential strength to credence, prescribe maximal credence for *Dog*:

---

[4] For classic discussion of the issue see Ian Hacking's "Slightly More Realistic Personal Probability" in *Philosophy of Science* 1967, and for state-of-the-art discussion see Richard Pettigrew's "Logical Ignorance and Logical Learning" *Synthese* 2021.

[5] For classic discussion of the issue see David Lewis' "A Subjectivist's Guide to Objective Chance", in (Jeffrey, R. eds.) *Studies in Inductive Logic and Probability* (University of California Press 1980) or Vann McGee's "Learning the Impossible" in Eells and Skyrms (eds.) *Probability and Conditionals: Belief Revision and Decision* (CUP 1994), and for state-of-the-art discussion see Kenny Easwaran's "Regularity and Hyperreal Credences" *Philosophical Review* 2014.

$$[(1)^{+} \,\&\, (2)^{+} \,\&\, (3)^{+}] \;\Rightarrow\; \mathrm{cr}(\mathit{Dog}) \cong 1.$$

This means our formalism treats *Dog* on a par with ☺: assigning both maximal credence. Since ☺ enjoys no epistemic risk of error, while *Dog* may well be false, a natural thought is that this result is unacceptable. The palpable asymmetry between ☺ and *Dog* suggests that a model of rationality is acceptable only if it marks the asymmetry somehow. Our formalism looks to fail this test since it characterizes a person's take on a claim with credence.[6]

We can now see that any model endorsing $(1)^{+}$, $(2)^{+}$, and $(3)^{+}$ faces difficulties. Those difficulties involve treating claims with extreme credence when such credence seems inappropriate. Let us try to understand in each case why that problem occurs.

4. Diagnosing the Logic Case.

Why does our formalism seem to go wrong in the logic case? Did we err with our starting assumptions about rationality? Did we err in our formalization of those assumptions? How do we diagnose the case?

Recall that *Log* is a truth of logic which no human can easily understand (much less prove). For this reason, it looks as if an acceptable model of rationality will treat *Log* differently than a trivial truth of logic. Yet the probability theory of our formalism obliges maximal probability for all truths of logic. This together with our model's conversion of evidential probability to rational credence forces the view that maximal credence is obliged for all truths of logic. Hence our model treats all truths of logic on a par with trivial ones like ☺.

If appearances are on the right track here—if truths like *Log* and ☺ should not be treated the same way—it is the mathematics of our formalism which makes for a mismatch between model and fact. Since the mathematics insists on maximal probability for both sorts of truths, it is responsible for difficulty in the logic case. This does not mean probability theory in our model is mistaken *as* a bit of mathematics, of course. It just means that probability theory in our model is over-eager in a certain way when applied to rationality. We may thus reach a conditional verdict: *if Log* and ☺ should be treated differently in the formal theory of rationality, some sort of non-standard mathematics should be used.

A natural line of thought rejects the if-part of this verdict. It is based on the idea that models of rationality incorporating $(1)^{+}$, $(2)^{+}$, and $(3)^{+}$ target ideal rationality rather than its everyday cousin. The perspective maintains that ideal rationality requires *Log* and ☺ be treated the same way, since ideal agents see through complexities which lead humans to treat *Log* and ☺ differently. The fact that humans do not—and perhaps even cannot—lend credence in line with our formalism shows nothing about whether *Log* and ☺ should be treated the same way in the ideal. Since both claims are truths of logic, they should both be treated the same way. To think otherwise is to conflate ideal and everyday rationality.

There seems a lot right in this reaction to the logic case. It *is* possible to distinguish ideal from everyday rationality, after all, just as it is possible to distinguish ideal from everyday

[6] This situation turns in no way on the specific claim offered by your evidential source—any claim would do. No matter which claim is put forward our formalism ensures that credence for that piece of evidential information should be maximal. That looks plainly wrong. For classic discussion see Hartry Field's "A Note on Jeffrey Conditionalization" *Philosophy of Science* 1978, and for the state-of-the-art see Jason Konek's "The Art of Learning", *Oxford Studies in Epistemology* 2017.

morality, ideal from everyday decision-making, and so on. On every plausible way of spelling out the epistemic distinction, moreover, ideal and everyday phenomena differ in one crucial respect: the ideal involves fully worked-out phenomena while the everyday involves less-than-fully-worked-out phenomena. Applied to the rationality of opinion the point is that ideal opinion cannot be made better by reflection, attention, suppositional reasoning, or other methods of quotidian epistemic improvement, for ideal opinion is already worked out to the maximum degree. It is complete in all the ways promoted by everyday devices of rational improvement. Yet reflection, attention, and reasoning are the meat-and-potatoes of everyday rationality. Their deployment is central to the production of better and better everyday opinion.

When it comes to rationality *tout court*, then, full-dress theory should recognize an important distinction between an ideal and an everyday variety. Once this is done it is possible to argue with force that *Log* requires maximal credence for ideal rationality but not for everyday rationality, and, therefore, that ☺ requires maximal credence for both types of rationality. This line of argument makes for a diagnosis of what goes wrong with an intuition to the effect that our formalism misfires in the logic case. The thought is that intuition mixes-up ideal and everyday rationality. *Log* and ☺ differ when it comes to the latter but not when it comes to the former. Whether our formalism misfires in the logic case thus depends on which type of rationality is in play. If it is a variety which ordinary humans may hope to manifest in their everyday life, our formalism misfires badly in obliging maximal credence for *Log*. If it is a variety had only by ideal agents, our formalism performs admirably in the case.

Now, it is easy to note a distinction between ideal and everyday rationality. It is not so easy to develop a good understanding of the distinction. One reason for this is simple: doing the latter task requires answering a which-comes-first sort of question, namely, does ideal rationality come before its everyday counterpart, explanatorily, does everyday rationality come before its ideal counterpart, explanatorily, or is neither type of rationality more fundamental than the other?

Although super-important this question is not discussed a great deal in the epistemic literature.[7] But the relevant topic can be made clear by an analogy with the theory of political justice, since the analogue issue is both intuitive and has been extensively discussed in the literature on justice. So pretend for a moment that you are a political activist wishing to make your community more just. How should you proceed with your work?

One school of thought says that you should first get clear on the conditions needed for ideal justice. The relevant thought is that you should use information about ideal justice to improve your sub-optimal community by taking steps to make it better approximate the ideal. On this approach, political perfection is the literal lodestar of activism in a sub-optimal world. Ideal justice is the bedrock of everyday political improvement. The perspective entails that ideal justice is more basic than its everyday counterpart, since it is something which functions as a substantive target (and hence constitutive element) of that everyday counterpart.[8]

An alternative school of thought says that you should first assess how available options improve your community in justice-related ways, where the improvements are ***not*** understood by appeal to political perfection. The relevant thought here is that you should opt for a path forward by choosing the locally best of the available options. On this view, political perfection plays no role in the active improvement of your community—only everyday aspects of imperfect scenarios are relevant to the push for justice.[9]

---

[7] For brief discussion see Julia Staffel's *Unsettled Thoughts: A Theory of Degrees of Rationality* (OUP 2019).

[8] John Rawls thinks of things this way in *A Theory of Justice* (Harvard 1971).

[9] Amartya Sen thinks of things this way in *The Idea of Justice* (Harvard 2009).

These approaches differ in the role idealisation is given in well-grounded political activism. The first sees perfection as essential to the work, the second sees it as incidental to the work (at best). The first uses ideal justice to ground its everyday counterpart, the second puts ideal justice to the side and deals directly with less-than-ideal improvement of the imperfect world. These are ideal-first and everyday-first approaches to political activism. They correspond to a fundamental division in the theoretical area.

There is an important but under-discussed analogue division in the theory of epistemic rationality. To see how it works, suppose you are an epistemic activist rather than a political one: you realize—as Descartes did in the *Meditations*—that your beliefs, states of confidence, shifts of opinion, etc., have been sub-optimal throughout your life, and you wish to make them epistemically better. How should you proceed?

One school of thought says you should first get clear on conditions which make for ideal rationality. The relevant idea is that you should use information about those conditions to improve your worldview by taking steps to make it better approximate that of an ideally rational agent. If this is so, epistemic perfection is the lodestar of epistemic improvement in a sub-optimal agent. Ideal rationality is the bedrock of its everyday cousin. The perspective entails that ideal rationality is explanatorily more fundamental than its everyday cousin, for ideal rationality is the literal target at which everyday rationality aims.

An alternative school of thought says that you should directly assess available options for improvement in your epistemic situation; then you should update your worldview by opting for the locally best of the available options. On this view, epistemic perfection plays no role in the rational improvement of mind. Only everyday aspects of sub-optimal opinion is relevant to epistemic improvement.

These approaches differ in the role they give idealisation in well-grounded epistemic activity. The first sees epistemic perfection as a driving force in the everyday improvement of mind. The second sees it as incidental to such improvement (at best). The first is an ideal-first approach to everyday rationality, the second is an everyday-first approach. The former maintains that opinions are rational in an everyday way through similarity to ideally rational opinion. The latter maintains that opinions are rational in an everyday way through manifesting local features promoted by quotidian devices of epistemic improvement—inference, attention, double-checking, and so on—devices not individuated by appeal to the epistemic ideal.

Interestingly, both approaches are well known in the literature on rational belief. But only the ideal-first approach can be found in literature on rational credence. There is a hole in the literature on the everyday rationality of credence:

		Rôle of Idealisation	
		Ideal-1st	Everyday-1st
Type of Attitude	Belief	Belief-revision Theory, Default Logic, Non-monotonic Logic.	Theory of Defeasible Reasoning
	Credence	Theories of Bounded Rationality, Distance-from-the-ideal approaches	**???**

The bottom-right corner of this table should be brimming with theory, and that theory should be linked to other aspects of rationality. Such an approach would start with features promoted in the everyday improvement of credence—none of which are understood by appeal to ideally

rational credence—and it would use those features to specify *ceteris paribus* norms for the promotion of sensible credence. This would be the confidence-theoretic analogue of the theory of default logic or defeasible reasoning. There is no such approach in the epistemic literature.[10]

Reflection on the logic case thus exposes a hole in that literature. To fill it we need to understand everyday rationality of credence directly, in its own terms. This sort of view will differ from our take on ideal rationality in at least two respects: it will involve defeasible (or *ceteris paribus*) norms, and it will involve local rather than global norms.

5. Diagnosing the Coin Case.

Why does our formalism seem to go wrong in the coin case? Did we err in the starting assumptions about rationality built into that formalism? Did we err in our formalization of those assumptions?

Recall that in the coin case a fair coin is flipped an infinite number of times, and *Hog* is the claim that the coin lands heads-on-the-ground every time. Whether *Hog* is true is a contingent matter: everyone can see that it might be true and might be false. The problem is that our formalism entails *Hog* receives zero evidential probability, and, thereby, a prescription for zero credence. The formalism treats *Hog* as it treats the negation of our risk-free claim ☺. This looks wrong, of course, since everyone can see that ¬☺ has no chance of being true, while *Hog* has as much chance of being true as whatever turns out to be the case with the coin. Credal values in our formalism do not mark this important epistemic asymmetry.

The difficulty cannot be explained away by appeal to a mix-up between ideal and everyday rationality. After all, everyone can see the truth-value of *Hog* is a contingent matter, and everyone can see the truth-value of ¬☺ isn't like that (by stipulation). This means ideal and everyday agents should both treat *Hog* differently than ¬☺. Our model treats them the same way: the claims receive an evidential probability of 0, which converts to a prescription for minimal credence. Since credence is the model's mark if an agent's take on a claim, *Hog* and ¬☺ end-up being treated the same way. The difficulty cannot be avoided by appeal to a distinction between ideal and everyday rationality.

A common way to handle the problem here involves locating new mental states of some kind—mental states over and above real-valued credence—and using a formal representation of them to mark the palpable difference between *Hog* and ¬☺. This is done in various ways in the literature. For example:

- Some use non-zero infinitesimals in their model. These are non-zero numbers smaller than any standard real number. Epistemologists who use them give logical falsehoods zero probability, claims like *Hog* infinitesimal-but-non-zero probability, and convert numbers straightforwardly into prescription for credence. This involves new psychology

[10] The claim here is ***not*** that there is no work on the everyday rationality of credence. The claim is that there is no everyday-first work on the everyday rationality of credence. There is plenty of ideal-first work on the everyday rationality of credence, in both psychology and philosophy. For a good example of the former see Mike Oaksford and Nick Chater's ***Bayesian Rationality*** (OUP 2007), and for a good example of the latter see Paul Weirich's ***Realistic Decision Theory*** (OUP 2004).

marked by new mathematics, since infinitesimal credence is no part of the standard conception of credence.[11]

- Others use primitive conditional probability to mark a new kind of psychological state, basic conditional credence. They give logical falsehoods one role with respect to the new conditional credence and *Hog* another role. Here again new psychology is marked by new mathematics, with the coin problem being solved through new resources. This time the new resources involve a new type of conditional credence rather than a new type of unconditional credence, but in other respects the move is like the one sketched in the previous bullet.[12]
- Still others use primitive mathematical comparisons to mark a new kind of comparative confidence, psychologically basic comparative confidence. They give logical falsehoods one role with respect comparative tools and *Hog* another. New psychology is again marked with new mathematics and the coin problem is solved with the new resources. For our purposes this move is like the previous two in its essentials.[13]
- Some locate new psychology to solve the *Hog* problem but use aspects of our *old* formalism mark it. The idea is that trivial falsehoods relate one way to the new psychology and *Hog* relates another way, with these relations being tracked already in our formalism. The key thought is that we didn't interpret the formalism correctly before, since we didn't realise that part of it was tracking something psychologically important. We have new psychology here in the old mathematics, a new take on extant technical resources.[14]

All these reactions to the coin case make use of new psychological resources, attacking the problem with something over and above real-valued credence. Most relate their new psychology to new mathematics, some use old bits of the formalism instead. Either way the crux move is to insist that there is more to rational psychology than real-valued credence.

It can be hard to spot this lesson if we are dealing with ideal rationality. After all, that sort of rationality requires relevant mental states to exist and to align properly. With the theory of everyday rationality, though, the opposite is the case. Ordinary agents are sub-optimal precisely because their mental states are not as they should be—either they don't exist when they should, or they exist in some sort of misaligned way. This makes for sub-optimality in a way that never happens to ideal agents. Hence the theory of ideal rationality can place very much less emphasis on psychology than its everyday cousin.[15]

---

[11] See David Lewis' "A Subjectivist's Guide to Objective Chance", in Richard Jeffrey's *Studies in Inductive Logic and Probability* (Berkely 1980), or Brian Skyrms' *Causal Necessity* (Yale 1980). For the background mathematics see J.L.Bell's *A Primer of Infinitesimal Analysis* (CUP 1998).

[12] See Karl Popper's *The Logic of Scientific Discovery* (Hutchison & Co 1959), or Alfred Renyi 'On a New Axiomatic Theory of Probability' *Acta Mathematica Academiae Scientiarum Hungarica* 1955.

[13] See James Hawthorne's "A Logic of Comparative Support: Qualitative Conditional Probability Relations Representable by Popper Functions, in Alan Hajek and Christopher Hitchcock's *Oxford Handbook of Probability and Philosophy* (OUP 2016) or Harrold Jeffries' *Theories of Probability* (OUP 1939).

[14] For a move like this see Richard Jeffrey's reaction [in *The Logic of Decision* (Chicago 1965)] to Karl Popper's suggestion [in *The Logic of Scientific Discovery* (Hutchinson & Co 1959)] that multiple probability functions should be used to represent the opinions of an agent, or Kenny Easwaran's treatment of *Hog*-like claims in "Regularity and Hyperreal Credences *Philosophical Review* 2014.

[15] To see this *in extremo* see Tim Maudlin's "Credence—and Chance—Without Numbers (and Without the Euclidean Property)", http://philsci-archive.pitt.edu/19237/.

To see the point clearly consider two different formalisms. One takes ordinary probability as mathematically basic and defines conditional probability by appeal to it (in line with orthodox Bayesianism), the other takes conditional probability as mathematically basic and defines unconditional probability by appeal to it (in line with newer versions of the approach). The two theories are quite different when applied to the mind. The first holds fast to the suggestion—made by its formalism when taken at face value—that credence is psychologically basic and conditional credence is derivative. The second holds fast to the suggestion—made by its formalism when taken at face value—that conditional credence is psychologically basic and credence is derivative. These are incompatible views of conditional and unconditional credence. Doesn't that ***matter*** to the theory of rationality?

When ideal rationality is our target, the answer is "not so much". After all, ideal rationality obliges conditional and unconditional credence to exist and align in the right sort of way. Ideal rationality requires, for any claim A and any claim C, that strength of credence lent to C given A is well-measured by the result of dividing strength of credence lent to (A&C) by strength of credence lent to A. Here we have the familiar ratio

$$\mathrm{cr}(C/A) \;=\; \frac{\mathrm{cr}(A\&C)}{\mathrm{cr}(A).}$$

When credence for A is non-zero, the equation holds on both approaches to probability mentioned earlier. From a value-of-probability perspective, therefore, it does not matter if we hear the schema reductively from left-to-right, reductively from right-to-left, or even non-reductively`. What matters is that our formalism can be used to track epistemic differences between *Hog* and trivial falsehoods of logic.

If we turn our attention to everyday rationality, however, differences of mathematical interpretation become much more important. The theory of everyday rationality must be sensitive—both conceptually and in its formalism—to explanatory relations between types of psychological state. One reason this is true is because phenomena modelled by theory are fragmentary in nature: opinions are not fully worked out, shifts of opinion are local rather than global, and so forth. We need a formalism to reflect when a person works out a better-but-still-imperfect take on things, when an agent chases down some-but-not-all consequences of what they accept, when they lend credence to things for the first time, and more. This is how rationality works in the real world. Formal theories of everyday rationality must use mathematics to reflect the fragmentary character of everyday opinion. Formal approaches to ideal rationality may rise above this need, but it is the bread-and-butter of everyday rationality.

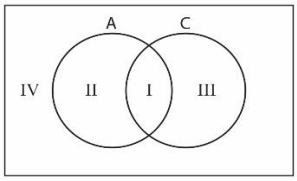
In probability theory it is normal for conditional probability to reduce to its unconditional cousin, though sometimes the reverse reduction is done (as we've seen). Each approach functions well for mathematical purposes, but neither should be used in a formal theory of everyday rationality. Taken at face value, after all, each misleads ***badly*** about the target psychology. With thinkers like us, conditional credence does not reduce to credence, and credence does not reduce to conditional credence. Our psychology manifests two types of self-standing attitude: one is a graded two-place relation between a lender of credence and a content to which credence is lent, the other is a graded three-place relation between a lender of conditional credence and a pair of contents (one of which receives grades of credit while the other is taken conditionally to be true). This is how things work for us psychologically. Our

approach to everyday rationality should mark it explicitly. If we're doing formal epistemology, the mark should be an explicit part of our formalism.

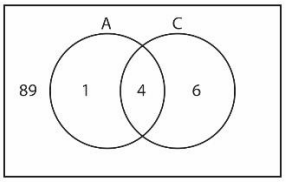
Consider the fact that we sometimes use rational credence to fix rational conditional credence, and other times we do the reverse. Here is a case of the first sort

DROP ZONE

You are one of 100 commandos about to be dropped into battle. The drop zone looks like a four-celled partition of logical space, with undivided cells labelled to match lines on a truth-table:



You have full rational credence in the view that your commanding officer knows the situation. She tells you that the C-section of the drop zone leads to ***C***ertain death, but elsewhere in the drop zone is safe. She comments that commandos will be distributed randomly as follows:



You set credence accordingly. Then another commando asks how likely is it that someone will die if they land in the A-section of the drop zone?[16]

Although you are well placed to answer this question, you'll need to *work out* the answer. When you work out the answer, however, there will be a stage in your thinking when credence for (A&C) will be 4% and credence for A will be 5% yet no credence for C given A will exist. It will be by virtue of the first two credal states that you will come to an 80% credence for C given A.

When orthodox formalism is applied to the mind, and taken at face value, conditional credence reduces to ratios of unconditional credence. When answering the question above there is a stage in your thinking when the ratio associated with conditional credence for C given A exists—since credence for (A&C) and credence for A both exist—yet conditional credence for C given A does not exist. Creating such a conditional credence is the whole point of your

[16] See the Judy Benjamin case in Bas van Fraassen's "A Problem for Relative Information Minimizers in Probability Kinematics" *British Journal for the Philosophy of Science* 1981, or chapter 4 of *The Rational Mind* (OUP 2020).

reckoning. This shows that orthodox probability should not be naively applied to the mind. Doing so results in a mistaken conception of conditional credence.

Now consider a case to illustrate that a reverse reduction is likewise mistaken:

THE EXAM

> The Exam Oracle says "I will choose one of five exam topics at random. Three are easy and two are hard. There's a 90% chance that Fritz will pass if an easy topic is picked, and a 30% chance that Fritz will pass if a hard topic is picked." You set credence and conditional credence accordingly. Then the Oracle asks: "how likely is it that Fritz will pass the exam?".[17]

Although you are well placed to answer this question, once again you'll need to work out the answer.

Here is one way to do it. You start out 60% sure that Fritz will face an easy exam and 40% sure that he will face a hard exam. You start out 90% sure that Fritz will pass the exam given he gets an easy topic, and 30% sure that he will pass given he gets a hard topic. These two states of mind rationally lead you to be 54% certain that Fritz will pass an easy exam (since that is 60% of 90%) and 12% certain that he will pass a hard one (since that is 30% of 90%). You are sure that Fritz will pass the exam exactly if he passes either an easy exam or a hard one, so you become 66% sure that he will pass the exam. This you report by saying that it is 66% likely that Fritz will pass the exam.

This is a bit of everyday reckoning. It involves fixing credence for claims by appeal to conditional credence. You pin down a credence for (Pass & Easy) by appeal to credence for Easy together with conditional credence for Pass-given-Easy, and you pin down a credence for (Pass & Hard) by appeal to credence for Hard together with conditional credence for Pass-given-Hard. You make use of a general fact about probability, namely, that for any claims Φ and Ψ

$$\text{pr}(\Phi\&\Psi) \;=\; \text{pr}(\Phi) \times \text{pr}(\Psi/\Phi).$$

This general fact is a trivial consequence of the ratio formula discussed earlier. It is central to everyday reckoning in that rational mental states corresponding to its right-hand side very often help to fix rational mental states corresponding to its left-hand side. In everyday life, conditional credence helps to pin down its unconditional cousin.

DROP ZONE shows that credence can be used rationally to fix conditional credence. **THE EXAM** shows that conditional credence can be used rationally to fix credence. Both sorts of moves are canonical within everyday rationality. Neither is relevant to ideal rationality. Formalism needed for an effective model of the latter need not reflect the fragmentary nature of everyday cognition. It is crucial to a formal theory of everyday rationality that this sort of reflection take place. Yet conceptual priority in a formalism should echo explanatory priority in psychological targets, since reflecting the nature of cognition is the *raison d'être* of formal work in the field. The deep lesson of the coin case, then, is that it is best in epistemology to be crystal clear on basic mental phenomena. This is particularly important for the theory of everyday rationality, since that sort of rationality turns on partial or fragmentary opinion.

---

[17] See Dorothy Edgington's "On Conditionals" *Mind* 1995.

6. Diagnosing the Dog Case.

Why does our formalism seem to go wrong in the dog case? Did we err in our starting assumptions about rationality? Did we formalize those assumptions in a bad way? What is our diagnosis of the situation?

In the dog case an evidential source plumps for the view that there is a dog before you. Perhaps it looks as if there is a dog before you, or it sounds as if there is a dog before you, or a trusted source maintains that there is a dog is before you. By stipulation *Dog* is the claim plumped for by an evidential source and hence part of the total evidence. This means *Dog* enjoys maximal evidential probability on that evidence, and, thereby, maximal credence is called for. Whatever evidential sources turn out to be, though, it hard to believe their use requires absolute certainty. Hence our model looks to misfire in the dog case, pushing for extreme credence when something more moderate seems called for.

In this respect the dog case is like the other two before us. But notice an important asymmetry in the three cases: the logic and coin cases turn on two bits of theory, the dog case requires three assumptions about rationality. Recall the elements in play:

$(1)^+$ $\quad I \in \mathbf{TE}$.
$(2)^+$ $\quad \underline{\text{ES}}(h,\mathbf{TE}) = \underline{\text{Ev-Prob}}(h/\mathbf{TE})$
$(3)^+$ $\quad \text{cr}(h) \cong \underline{\text{Ev-Prob}}(h/\mathbf{TE})$.

$(1)^+$ says that information got from an evidential source is part of the total evidence. $(2)^+$ says that evidential support for a claim on total evidence equals its evidential probability on that evidence. $(3)^+$ says that credence should match evidential probability on total evidence. In the logic case, we have the second and third bit of theory obliging maximal credence for *Log*:

$$[(2)^+ \,\&\, (3)^+] \;\Rightarrow\; \text{cr}(\mathit{Log}) \cong 1.$$

In the coin case, we have the same bits of theory obliging minimal credence for *Hog*:

$$[(2)^+ \,\&\, (3)^+] \;\Rightarrow\; \text{cr}(\mathit{Hog}) \cong 0.$$

In both cases faulty prescription results from our starting assumption about evidential support together with our starting assumption about its relation to rational credence. The dog case turns on an extra bit of theory put with these assumptions:

$$[(1)^+ \,\&\, (2)^+ \,\&\, (3)^+] \;\Rightarrow\; \text{cr}(\mathit{Dog}) \cong 1.$$

Evidential input is essential to the faulty prescription here. It is inessential to difficulties seen in the first two cases.

The dog case plagues any view on which our three starting assumptions about rationality hold. If evidence consists in bits of information, bits of information enjoy levels of support on the total evidence, and credence should match those levels of support, then, credence for evidence should be maximal. Since credence for evidence should not always be maximal, views which start out like ours face a credence-for-evidence problem.

Or so it seems.

One might agree with our starting assumptions but deny that that they lead to a problem. After all, we have seen that our starting assumptions can be defended by appeal to ideal rationality. Perhaps they can also be defended by appeal to ideal sources of news. Such sources would be *infallible*, after all, always serving-up nothing but correct information. For this reason, it might be thought that rational reaction to their output involves only maximal credence. If that is so, however, there is no credence-for-evidence problem at all. There is only apt formalization of rational reaction to ideal news.

This line of thought closely mimics our diagnosis of the logic case. Just as a defender of our assumptions about rationality can insist that the appearance of misfire in the logic case results from a mix-up of ideal and everyday rationality, so a defender of those assumptions might insist that the appearance of misfire in the dog case results from a similar mix-up. This time the mix-up would not be one involving ideal and everyday rationality but instead would involve infallible and fallible news. Once it is clear that our model targets agents with infallible input, the thought goes, the appearance of misfire in the dog case evaporates.

My initial reaction to this line of thought was unsympathetic, but I now think there is mileage in it. After all, whatever sources of evidence turn out to be—mechanisms generating perceptual states, states of supposition, states of testimonial information—sources of evidence are bits of epistemic equipment. They are part of the machinery which generates rational states of mind and rational transitions between states of mind. This means sources of evidence can be idealized like any other piece of epistemic equipment. Just as inferential mechanisms can be idealized so that one ends-up theorizing about single-step shifts in an entire worldview—grand kinematic moves *à la* Conditionalization or Jeffrey's rule[18]—so it is that news-gathering sources can be idealized so as to serve-up truth and only truth.

This is a non-trivial idealization, to be sure, but it is one with reputable defenders in the literature—even when everyday news-gathering sources are discussed. When those sources put forward perceptual information, for example, the relevant idealization individuates them so that they are only at work when an agent is perceiving the world aright (i.e. only when an agent is in the so-called Good Case). Similarly: when evidential sources plump for testimonial information, the relevant idealization individuates them so that they are only at work when knowledge-based testimony is involved. And when evidential sources put forward memorial information, the relevant idealisation individuates them so that they only make available past acquaintance with fact. And so on.[19] These idealizations have one thing in common: they individuate news-

[18] See chapter 11 of Richard Jeffrey's *The Logic of Decision* (University of Chicago Press 1965) or chapter 4 of *The Rational Mind*. As the latter notes, neither Conditionalization nor Jeffrey's rule is well viewed as a prescription for contact-point reaction in attitude to news. Instead, they are best viewed as rules for coherence recovery after coherence has been lost due to such reaction to news. When Jeffrey proposed his generalization of Conditionalization, work in the area was routinely mal-shaped by the dogma that only a belief can be reason for belief. In that context it is possible to squint hard and see Conditionalization and Jeffrey's rule as prescription for contact-point reaction in attitude to news. But those days are long gone. Modern epistemology follows common-sense and rejects the idea that shift of belief counts as receipt of news from an evidential source (at least in ideal agents). This allows epistemic theory to side-step the isolation objection (also known as the frozen-attitudes objection)—see Alvin Plantinga's *Warrant: The Current Debate* (OUP 1993). But there are in fact many reasons to follow common-sense here, many reasons to insist that receipt of evidential information is one thing and contact-point reaction in attitude to such receipt is another (at least for ideal agents). Once that is done, though, neither Conditionalization nor Jeffrey's rule even speak to the credence-for-evidence problem (much less solve it).

[19] John McDowell defends a view a like this for perceptual mechanisms in *Mind and World* (Harvard 1994). Mikkel Gerken defends a view like this for testimonial mechanisms in "Internalism and Externalism in the Epistemology of

gathering sources so that they only deliver the epistemic goods. On such an approach, sources of evidence always anchor an agent to facts. If evidential sources are idealized this way, rational reaction to news might well involve maximal credence.[20]

It is an empirical issue whether humans have fallible or infallible sources of news. Like most I incline to the view that we have only fallible sources of news. But we need not decide the issue here, for it is our job as philosophers to clarify how rational credence should play out in both sorts of case—when agents have fallible sources of news and when they have infallible sources of news.[21] The key question each time is this: how should credence for evidence be and why? Full-dress theory must answer this question for agents with infallible sources of evidence. It must also do so for those—like us, say I—who suffer from only fallible sources of news.

We assume without argument that models of rationality built on $(1)^+$, $(2)^+$ and $(3)^+$ give acceptable verdicts for infallible sources of evidence. When the incoming news cannot be mistaken, the assumption is that maximal credence for evidence is called for. The *tough* issue centres on credence for evidence when sources of news are fallible, when they make mistakes. Assume in what follows that our sources of evidence are fallible in this way. Then the issue is whether credence for evidence should be maximal in that case.

If the answer is Yes, our three starting assumptions about rationality are unthreatened by the fallibility of the news. If the answer is No, at least one of our starting assumptions is false: either evidence does not come in bits of information, information does not always enjoy evidential standing in light of total evidence, or credence for evidence should not always match a piece of information's standing on total evidence. Since we're assuming that the first of these options is not so, it follows that one of two things must be true. Either

(A) Bits of evidence are not always maximally supported by the total evidence,

or

(B) Credence for evidence should not always match evidential support on the total evidence.

Which option is best?

Well, a line of thought favoring (A) makes use of an analogy with question-begging arguments. To see how it works, consider a schematic argument of this kind:

(QBA)

P_1
P_2
•
•
P_n

Testimony", *Philosophy and Phenomenological Research* 2013. M.G.F.Martin defends such a view for memorial mechanism in "Out of the Past: Episodic Recall as Retained Acquaintance", in Christoph Hoerl and Teresa McCormack's *Time and Memory: Issues in Philosophy and Psychology* (OUP 2001). Timothy Williamson generalizes the Oxonian perspective in *Knowledge and Its Limits* (OUP 2000).

[20] But can't agents with such idealised news-gathering mechanisms receive news that their evidential sources are fallible? And shouldn't they have less-than-maximal credence for evidence in that case? No: this isn't possible, for ideal sources of evidence are infallible. They cannot suggest their own fallibility, for that would be a mistaken suggestion. Idealised sources of news never make such a suggestion.

[21] We leave the combination view as a homework problem.

C

∴ C

Any instance of this schema will be an argument with a conclusion which is also one of its premises. Any argument like that begs the question when it comes to its conclusion. In all normal senses of logical support, though, premises of **(QBA)**-style arguments maximally support their conclusions. The issue is whether they *establish* those conclusion as well as logically support them.

This depends on what you mean by "establish". The premises of **(QBA)**-style arguments do tell stories the truth of which guarantees the truth of their conclusions. Guarantees of that sort have proven useful in logic, mathematics, computer science, and other areas of concern. In a broad and useful sense of "establish", therefore, premises of **(QBA)**-style arguments establish their conclusions. But such premises do not—and indeed arguably cannot—underwrite rational transitions *from* an intellectual place where a conclusion's truth-value is an open question *to* one where it is no longer an open question. Since premises of a **(QBA)**-style argument explicitly contain the argument's conclusion, commitment to the argument's premises involves commitment to the argument's conclusion. This is why rational endorsement of the conclusion cannot be generated by commitment to the argument's premises. Despite those premises establishing the argument's conclusion—in a broad and useful sense of "establish"—the premises of a **(QBA)**-style argument cannot create rational commitment to its conclusion.

When the topic is whether a given claim C is true, therefore, explicitly question-begging arguments for C are a hopeless tool in the formation of opinion. They cannot be used to shift rational opinion about whether C is true. One might suppose, though, that a signature role of evidence is to create new opinion about how a given subject matter turns out (or to give new support to an old opinion). If evidential claims work like premises of an argument, though, they are incapable of playing this role for their own alethic status. This suggests that no claim can be part of the evidence concerning its own truth-value. Hence a proprietary role of evidence suggests that total evidence somehow shifts from topic to topic to preclude explicitly question-begging situations. In the event, it is not automatically true that bits of evidence are maximally supported by the total evidence, for it is not automatically true that bits of evidence are part of the total evidence when it comes to their own truth-value.

This line of thought supports an (A)-style reaction to the credence-for-evidence problem. The reaction denies that bits of evidence are automatically supported by the total evidence. Since the notion of evidence varies from subject-matter to subject-matter, the notion of evidence ensures that no piece of information can be part of the total evidence relevant to its own truth-value. This creates a nifty reaction to the credence-for-evidence problem, as it removes the structure in our model which generates maximal support for the evidence.

Yet the move here creates theoretical questions of its own. Chief among them is the question of credence for evidence. Models based on $(1)^+$, $(2)^+$, and $(3)^+$ have the good-making feature of answering this question directly. They have the bad-making feature of answering the question mistakenly—at least when sources of news are fallible. If we grant that bits of evidence are not always maximally supported by the total evidence—and thereby reject $(2)^+$ in our model—we

avoid the prescription of maximal credence for fallibly-generated evidence. But we are left with the difficult question of what credence for evidence should be.[22]

Moreover, contextualism about evidence—of the sort we're contemplating, anyway—seems illicitly to run together propositional and doxastic rationality.[23] Since the former has to do with how evidential relations pan out between contents, and the latter has to do with how those relations should be exploited in the formation of opinion, the contextualist line under discussion in effect takes it as read that evidential relations are always to be exploited in the formation of opinion. That is an unguarded assumption at best or a conflation of propositional and doxastic rationality at worst. Either way the resulting line fails to refute the idea that information is maximally self-supporting. Instead it highlights the fact that self-support cannot be exploited in the formation of opinion.

Here's a simple way to make the point. Let **ES**(-,-) be an evidential support function and stipulate that

$$\mathbf{ES}(\alpha, \beta) = \underline{x}$$

codifies the $\underline{x}$-level of support that β gives α. Then we can say that **ES**(-,-) is ***self-addressed*** in situations like this:

$$\mathbf{ES}(\Phi, \Gamma) = \underline{x}, \qquad \text{with } \Phi \in \Gamma.$$

Since evidential support has to do with propositional rationality, which in turn has only to do with contents relating evidentially to contents, it follows that whenever $\Phi \in \Gamma$

$$\mathbf{ES}(\Phi, \Gamma) = \text{maximum}.$$

After all, every content does maximally well on its own truth, so self-addressed evidential support is automatically maximal.

This does not depend on evidential support manifesting the structure of a classic probability function, nor the structure of a Dempster-Shafer support function, nor the structure of any other formal tool in epistemology. The point turns solely on the fact that evidential support is maximal when self-addressed. The credence-for-evidence ***problem***, therefore, has nothing to do with the scope or limits of formal epistemology.[24] It turns solely on the fact that self-support fails automatically to translate into prescription for credence.

This point should look familiar. To see why, suppose $\Phi_1, \Phi_2, \ldots, \Phi_{large}$ are each bits of evidence got from a particular source S. Suppose you realize that S is the source for each Φ–claim and that the number of claims is large. Suppose also that you are good at creating new

[22] See Hartry Field's "A Note on Jeffrey Conditionalization" *Philosophy of Science* 1978, Dan Garber's "Field and Jeffrey Conditionalizaation" *Philosophy of Science* 1980, Jason Konek's "The Art of Learning" *Oxford Studies in Epistemology 2017*, Jim Pryor's "Problems for Credulism", in Chris Tucker's *Seemings and Justification: New Essays in Dogmatism and Phenomenal Conservatism* (OUP 2013), or Jonathan Weisberg's "Commutivity or Holism: A Dilemma for Conditionalizers" *British Journal for the Philosophy of Science* 2009.

[23] See Jonathan Kvanvig and Chris Menzel's "The Basic Notion of Justification" *Philosophical Studies* 1990.

[24] *Pace* Timothy Williamson's reply to Mark Kaplan in Patrick Greenough and Duncan Pritchard's *Williamson and Knowledge* (OUP 2009).

bits of evidence by inferring things from old bits of evidence. Then you will have maximal propositional support for the view that S has been fully accurate across a large range of cases:

ES(Large-#-of-fault-free-S-endorsements/your evidence) = maximum.

After all, the claim [S endorses Φ_1 when Φ_1], the claim [S endorses Φ_2 when Φ_2], etc., all the way down to the claim [S endorses Φ_{large} when Φ_{large}] will all be inference-based bits of evidence for you. But in this situation (and for this reason) should you lend maximal credence—or anything approaching maximal credence—to the view that S is faultless over a large range of cases? No, obviously not. Evidential support doesn't generate doxastic prescription in that way.

It is impermissible to adopt a view about the *bona fides* of a source of information based on evidence got largely from that source. This is (what I think of as) the ***Münchausen*** constraint on the adoption of attitudes. Flouting it amounts to yanking oneself out of an epistemic pit by pulling one's own ponytail (as Baron Münchausen yanked himself out of a pit while sitting on a horse). No such epistemic effort can properly ground a stance on the *bona fides* of an evidential source.

The Münchausen constraint plays a central role in the conversion of evidential support to credal strength. In the bootstrapping case just mentioned, you have maximal support for the view that endorsements of a particular source are accurate. That support derives almost exclusively from first-order views got from the source in question. By the Münchausen constraint, it is not legitimate to convert this content-based support into doxastic commitment about the *bona fides* of the relevant source. Conversions like that irrationally beg the question.

The deep lesson of the dog case, then, concerns credence for fallible evidence. It is clear that strength of credence for such evidence should not match strength of its support on the news. It is not clear what credence for such evidence should be. Just as the *Log* case indicates that we need an everyday-first approach to rational credence, and the *Hog* case indicates that we need to be clear in that theory of everyday rationality about fundamental psychology, so the *Dog* case indicates something important about our approach to sub-optimal epistemic situations, namely, that we need a story about rational credence for fallibly generated evidence. These are important lessons for everyday rationality. Each of them comes to the fore when epistemic models misfire.