Konek, J. (2016). Probabilistic Knowledge and Cognitive Ability. *Philosophical Review*, *125*(4), 509-587. https://doi.org/10.1215/00318108-3624754

Probabilistic Knowledge and Cognitive Ability

Jason Konek

Abstract

Moss (2013) argues that degrees of belief or credences can amount to knowledge in much the way that full beliefs can. This paper explores a new kind of objective Bayesianism designed to take us some way toward securing such knowledge-constituting credences, or 'probabilistic knowledge'. Whatever else it takes for an agent's credences to amount to knowledge, their success, or accuracy must be the product of *cognitive ability* or *skill*. The brand of Bayesianism developed here helps ensure this ability condition is satisfied. Cognitive ability, in turn, helps make credences valuable in other ways: it helps mitigate their dependence on epistemic luck, for example. What we end up with, at the end of the day, are credences that are particularly good candidates for constituting probabilistic knowledge. What's more, examining the character of these credences teaches us something important about what the pursuit of probabilistic knowledge demands from us. It does *not* demand that we give hypotheses equal *treatment*, by affording them equal credence. Rather, it demands that we give them equal *consideration*, by affording them an equal chance of being discovered.

Keywords: Probabilistic Knowledge · Cognitive Skill · Accuracy · Explanation

On the Bayesian view, belief is not an on-off matter. Opinions are *gradable*. An agent might be more confident that it will rain in London this afternoon than in Leeds, nearly certain that the exams she left on her desk yesterday did not grade themselves over the evening, etc.[1] And these opinions pin down different *truth-value estimates* for different propositions. (We follow de Finetti and Jeffrey in thinking of propositions as variables that take the value 1 at worlds where they are true, and 0 where false. Truth-value estimates are estimates of the value, 0 or 1, that the proposition takes at the actual world.) For example, if you are more confident that it will rain in London (call this proposition 'LONDON') than in Leeds (call this proposition 'LEEDS'), then your state of opinion 'pins down' a higher truth-value estimate for LONDON than LEEDS in the following sense: every assignment of truth-value estimates that rationalizes, or makes sense of your opinions attaches a higher value (a larger truth-value estimate) to LONDON than to LEEDS.[2]

When an agent's opinions about propositions $X_1, X_2, \ldots$ are *so* rich and specific that they pin down a *single* estimate x_i (a real number between 0 and 1 inclusive) of the truth-value of each X_i, we say that she has precise *degrees of belief* or *credences* for $X_1, X_2, \ldots$ Having a precise credence x_i for a proposition X_i, then, does not require having a full belief about the objective probability of X_i, or *explicitly* judging that x_i is the best estimate of X_i's truth-value, or anything of the sort. (Neither does it preclude this.) It simply requires having a doxastic state or state of opinion with a particular property, *viz.*, the property of being rationalizable only by sets of truth-value estimates which say: x_i is the best estimate of X_i's truth-value.[3]

[1] I follow many authors in using 'opinions' as a catch-all to refer any type of doxastic attitude whatsoever (*cf.*, Kaplan (2010) and Joyce (2010)). So, for example, full beliefs, comparative beliefs (judgments of the form *X is more likely than Y*), independence judgments, etc., all count as opinions, on this view.

[2] Kraft et al. (1959) and Scott (1964) provide representation theorems that specify exactly when an agent's comparative beliefs 'pin down' precise truth-value estimates for all comparable propositions, in the sense of singling out a unique set of truth-value estimates which rationalize (represent) those beliefs.

[3] Some Bayesians take credences to be *sui generis* doxastic attitudes: degreed beliefs with precise strengths (*cf.*, Pettigrew (2013a)). On this view, an agent who only has comparative beliefs, for example (opinions of the form *X is more likely than Y*) — a rather different type of doxastic attitude — does not count as having precise credences, regardless of how rich and specific those comparative beliefs happen to be. This contrasts starkly with the views of Savage (1954), Jeffrey (1965) and Krantz et al. (1971). The account of credences outlined here is more permissive. *Any* agent whose opinions are rationalizable by a unique set of truth-value estimates — whether she has *sui generis* doxastic attitudes of the sort Pettigrew envisions, or comparative

When an agent has precise credences, we call the function c that maps each proposition X_i to its corresponding truth-value estimate x_i her *credence function* or *credal distribution*.[4] When they exist, these credal values, $c(X_i) = x_i$ (the values assigned to $X_1, X_2, \ldots$ by her credence function), serve as numerical measures of how confident someone in the corresponding doxastic state can be said to be of the X_i, where $c(X_i) = 0$ and $c(X_i) = 1$ represent minimal and maximal confidence, respectively. So, for example, if you have a credence of 0.99 that the exams on your desk did not grade themselves over the evening, then 0.99 serves as a numerical measure of how confident you are in that proposition. You are ninety-nine times as confident that they are still sitting on your desk ungraded as not. We will engage in the useful fiction, from here on out, that the agents under consideration have perfectly precise credences.

Credences have a range of epistemically laudable properties, just like full beliefs. Just as full beliefs are evaluable on the basis of their *truth*, for example, credences are evaluable on the basis of their *accuracy*. An agent's credence in a proposition X is more accurate the more confidence she invests in X if X is true (*i.e.*, the closer $c(X)$ is to 1), and the less confidence she invests in X if X is false (*i.e.*, the closer $c(X)$ is to 0). Accuracy is a matter of getting close to the truth, in this sense.[5] Just as full beliefs capture more or less appropriate responses to the available evidence, so too do credences. If you are playing poker in a casino, you should (probably) have something like a credence of 0.0001 that you will be dealt a royal flush, in light of your evidence, rather than a credence of 0.9999 (unless you

beliefs, or some combination of the two — counts as having precise credences.

[4]Most authors treat 'credal distribution' as a slightly more general term than 'credence function', but use the two interchangeably when additional precision is unnecessary. We will follow suit. On this usage, an agent's credence *function* $c : \mathcal{F} \to [0,1]$ is defined on the full algebra $\mathcal{F}$ of propositions (closed under negation and disjunction) that she has opinions about. An agent's credal *distribution*, on the other hand, can refer to either (i) her total credence function c (*cf.*, Moss (2013) and Titelbaum (2015)), or (ii) the restriction of c to some contextually salient partition of interest, $\mathcal{X} = \{X_1, \ldots, X_n\} \subset \mathcal{F}$ (*cf.*, Joyce (2005, 2009)).

[5]We will make this characterization more precise in and §4.2 and §5.2 using *epistemic scoring rules* or *inaccuracy scores*. A scoring rule $\mathcal{I}$ maps credence functions c and worlds w (consistent truth-value assignments) to non-negative real numbers, $\mathcal{I}(c, w)$. $\mathcal{I}(c, w)$ measures how inaccurate c is if w is actual. Different scores capture different ways of valuing closeness to the truth.

know that the game is rigged, or something of the sort); the former is a more appropriate response to your evidence than the latter. And just as full beliefs are produced by more or less reliable mechanisms, so too are credences. Wishful thinking, for example, tends to produce fairly inaccurate credences, and so is a fairly unreliable credence-producing mechanism.

Sarah Moss (2013) argues that credences share even more in common with full beliefs than many Bayesian epistemologists have thought (*cf.* Moss 2013, p. 2). In particular, she argues that credences can constitute *knowledge* in much the way that full beliefs can.[6] Just as your belief that smoking causes cancer might constitute knowledge, so too might your extremely low credence that prayer prevents cancer. Moss calls this latter sort of knowledge 'probabilistic knowledge'. The aim of this paper is to develop a novel brand of objective Bayesianism — a theory which says, for any body of evidence, which credences to have given that evidence[7] — designed to take us some way toward securing probabilistic knowledge.

In §1, I rehearse a few reasons for countenancing probabilistic knowledge. In §2, I propose two necessary conditions on probabilistic knowledge — an anti-luck condition and an ability condition — and explore the relationship between the two. I argue that, for the purpose at hand, *i.e.*, designing helpful formal tools meant to help us secure probabilistic knowledge, we ought to focus on the *ability* condition. In §3-4, I search for a way to sift credal states that satisfy this ability condition from ones that do not. I argue that a particular 'summary statistic' of credal states can help; the smaller this statistic, the greater the extent to which its accuracy is a product of cognitive ability. In §5, I examine a popular

[6]Following Moss and others, I use 'constitutes' to refer to (as Moss puts it) the metaphysically innocuous relation that obtains between your doxastic and epistemic states (*cf.* Moss (2013), fn. 11, pp. 12-3). This is just the relation that holds when, *e.g.*, a full belief 'amounts to', or 'counts as', or 'rises to the level of' knowledge.

[7]More cautiously, objective Bayesian theories prescribe adopting certain 'prior' (pre-experiment) credences relative to certain types of 'prior' evidence, and perhaps contexts of inquiry or decision.

objective Bayesian principle for choosing prior (pre-experiment) credences in various contexts of inquiry: the maximum entropy principle, or MaxEnt. I use our summary statistic to evaluate whether MaxEnt delivers skillfully produced posterior (post-experiment) credences, and hence good candidates for probabilistic knowledge. In §6, I describe a novel objective Bayesian principle: the maximum sensitivity principle, or MaxSen. I argue that MaxSen yields better candidates for probabilistic knowledge than MaxEnt. It yields credences whose accuracy is, to the greatest extent possible, the product of cognitive ability. The upshot: MaxSen takes us some way toward securing probabilistic knowledge. In §7, I explore what MaxSen teaches us about the nature of cognitive ability and probabilistic knowledge. In §8, I summarise the preceding discussion. In §9, I address two pressing concerns. Finally, in §10, I discuss limitations of the MaxSen principle.

1 Probabilistic Knowledge

Suppose that Amy is waiting on a package. She calls the post office to find out whether it was delivered this morning. The receptionist checks the driver's morning itinerary and says:

(1) It might have been.

(2) More likely than not.

(3) It probably was.

To a first approximation, (1) calls for giving not-too-low credence to the proposition that her package was delivered (credence above some fairly minimal, contextually determined threshold, perhaps).[8] (2) calls for giving it more credence than its negation. (3) calls for

[8] See (Swanson, 2006, §2.2.2) for an account of 'might' along these lines.

high credence.[9] Suppose that Amy does what's called for. She responds in one of these ways. Then we might describe her new state of opinion as follows:

(5) Amy thinks that her package might have been delivered.

(6) Amy thinks that it's more likely than not that her package was delivered.

(7) Amy thinks that her package was probably delivered.

And if the receptionist's testimony is reliable, the following seem appropriate too:

(8) Amy knows that her package might have been delivered.

(9) Amy knows that it's more likely than not that her package was delivered.

(10) Amy knows that her package was probably delivered.

This raises a puzzle. On the one hand, Amy seems to acquire new knowledge. On the other hand, she seems to lack any full beliefs that might constitute this knowledge.[10] None of her *old* full beliefs (prior to updating on the receptionist's testimony) seem like good candidates. And (1)-(3) seem not to call for any *new* full beliefs. They obviously do not call for full belief that the package was *in fact* delivered. Neither do they call for any particular full belief about *chance*, *e.g.*, full belief that:

(11) There is a not-too-low chance it was delivered.

(12) There is a higher chance it was delivered than not.

[9] Assume that Amy's prior evidence is fairly typical: she has no special reason to think that the receptionist is being intentionally deceptive, or anything of the sort.

[10] By 'full belief' I mean the categorical doxastic attitude of accepting a proposition as true. Full beliefs are the theoretical loci of traditional epistemology: they are true or false, justified or unjustified, reliably produced or not, constitute knowledge or not, etc. Some epistemologists maintain that rational full beliefs are held with certainty, *e.g.*, Levi (1980). Many do not. Likewise, some epistemologists are eliminativists about full belief, *e.g.*, Jeffrey (1970). But many are not, *e.g.*, Kyburg (1961), Foley (1992) and Leitgeb (2013). I take no stand on these issues here.

(13) There is a fairly high chance it was delivered.

After all, Amy knows that the chance that the package was delivered earlier this morning is currently either 0 or 1.[11] And a chance of 0 is inconsistent with any of (11)-(13). So if the receptionist's testimony justified full belief in any of (11)-(13), then it would also justify (together with her background knowledge) full belief that the chance is 1! She could conclude that there is no chance — *zilch* — that her package was not delivered.

But her evidence justifies no such thing. For similar reasons, it does not justify any particular full belief about *epistemic probability*.[12] Hence the conundrum: Amy lacks any full beliefs that might constitute her new knowledge.

The way out of such conundrums, Moss suggests, is to countenance probabilistic knowledge: *credences can constitute knowledge, just as full beliefs can.*

Countenancing probabilistic knowledge not only allows us to make sense of run-of-the-mill knowledge ascriptions like (8)-(10). It also allows us to make sense of some puzzling ascriptions of *scientific knowledge* (of special interest for us). For example, consider the most recent assessment report (AR5) of the Intergovernmental Panel on Climate Change (IPPC). According to AR5, the majority of climate scientists think that:

(14) The Antarctic ice sheet might suffer abrupt and irreversible loss, even if anthro-

[11]Most accounts of chance satisfy what List and Pivato (2015) call *the chance-future desideratum*: the true chance distribution at any time assigns only 0 or 1 to past events. Notably, though, Meacham (2005, 2010) rejects this. See List and Pivato (2015) for further discussion.

[12]Epistemic probabilities, very roughly, provide a measure of the unique rational credence to have in a proposition relative to a particular body of evidence. To see that (1)-(3) do not call for full belief in a proposition about epistemic probability (if such probabilities exist), consider the following scenario. Amy has conclusive evidence about whether her package was delivered or not — perhaps a well-informed mathematician told her that it was delivered if and only if a particular mathematical proposition is true. And she knows that her evidence is conclusive. She knows that a perfectly rational agent with her evidence would have either credence 0 or 1 in the mathematical proposition, and accordingly have either credence 0 or 1 that her package was delivered. But Amy cannot tell which way her evidence points, so to speak, due to her uncertainty about the mathematical proposition. In that case, when the receptionist utters one of (1)-(3), she might reasonably use this information to weigh the two hypotheses about the the valence of her evidence, and arrive at some middling credence about whether her package was delivered. But she should not fully believe that the evidential probability takes some middling value. She knows it is either 0 or 1 (her evidence is conclusive). See (Eriksson and Hajek, 2007, pp. 206-207) for further discussion.

pogenic emissions of greenhouse gases are stopped. (Stocker et al., 2013, p. 18)

(15) Warming is more likely than not to exceed 4° C above pre-industrial levels by 2100 if we fail to take additional measures to curb greenhouse gas emissions. (*ibid*, p. 18)

(16) It is very likely that heat waves will occur more often and last longer as surface temperature rises over the 21*st* century. (*ibid.*, p. 10)

And if the IPCC's climate models are reliable, the following seem appropriate too:

(17) Those scientists know that the Antarctic ice sheet might suffer abrupt and irreversible loss, even if anthropogenic emissions of greenhouse gases are stopped.

(18) Those scientists know that warming is more likely than not to exceed 4° C above pre-industrial levels by 2100 if we fail to take additional measures to curb greenhouse gas emissions.

(19) Those scientists know that it is very likely that heat waves will occur more often and last longer as surface temperature rises over the 21*st* century.

(17)-(19) are puzzling for much the same reason that (8)-(10) are. Inquiry into the climate system typically leads researchers to adjust their *degrees of belief* or *credences* that various climatic events will or will not occur, as well as their credences about the chances of those events, the effects of different possible interventions, and so on. But it does not always yield decisive enough data to license new *full beliefs* about those matters. Nevertheless, such inquiry often seems to deliver new knowledge (the sort of knowledge described in (17)-(19)). Scientific inquiry seems to be a knowledge-producing enterprise, even if not always a full-belief-producing enterprise (as it often enough is not).[13] Countenancing probabilistic knowledge allows us to make sense of this fact.

[13] Of course, scientific inquiry often *does* lead researchers to adopt new full beliefs. And sometimes these full beliefs constitute knowledge. But even when inquiry only leads researchers to adjust their degrees of belief or credences, it still plausibly delivers new knowledge (evidenced by the appropriateness of (17)-(19)).

The probabilistic knowledge thesis — the thesis that credences can amount to knowledge, just as full beliefs can — has a range of other motivations too, Moss argues. Consider, for example, the following probabilistic Gettier case (adapted from Zagzebski (1996), pp. 285-6; see Moss (2013), p. 10 for a similar case):[14]

> Suppose that Mary has very good, but not perfect eyesight. She walks into the house with her daughter, glances across the living room at the man slouched in the chair by the fire, and whispers, "Quiet, honey. Daddy is probably sleeping." But Mary misidentifies the man in the chair. It is not her husband, Bob, but his similar-looking brother, Rob (whom Mary thought was out of the country). Luckily, though, Mary's high credence that Bob is sleeping in the living room is accurate. Bob, it turns out, is dozing in the *other* living room chair, along the far wall, just out of view.

In some sense, Mary has just the right credences. Her high credence that Bob is sleeping in the living room is *appropriate*, or *justified*, in light of her evidence, *viz.*, the appearance of a Bob-ish looking man slouched in the chair by the fire. And her high credence is fairly *accurate*, since Bob *is* in fact sleeping in the living room. (Recall, accuracy is a matter of getting close to the truth in the sense of lending high credence to truths and low credence to falsehoods.) Nevertheless, that credence is flawed, in some way. It is flawed in *just the same way*, it seems, that Smith's belief in the original Gettier case is flawed. Smith's belief that the man who will get the job has ten coins in his pocket is *appropriate*, or *justified*, in light of his evidence (reliable testimony that Jones will get the job, and a clear view of the contents of Jones' pockets). And Smith's belief is *true* (since Smith himself will get the job, and happens to have ten coins in his pocket). But Smith's belief, just like Mary's high credence, is true (accurate) primarily by luck.[15]

[14] See also Greco and Turri (2013), §6.

[15] Since Nozick (1981), epistemologists typically treat lucky true belief as true belief that violates some

The best explanation of the epistemic incorrectness in these two cases, Moss argues, is that Mary's high credence and Smith's belief both fail to constitute knowledge. One *could* attempt to explain the epistemic incorrectness in the two cases by positing that the absence of luck is a primitive epistemic virtue. But it would be better to identify some *positive* virtue that Mary's high credence and Smith's belief both lack, in just the way the probabilistic knowledge thesis does:

I. Constituting knowledge is an epistemic virtue.

II. Epistemic luck undermines knowledge.

III. Both Mary's high credence and Smith's belief are accurate primarily by luck.

C. Both Mary's high credence and Smith's belief fail to constitute knowledge, and so lack a certain (positive) epistemic virtue.

To summarise, the thesis that credences can amount to knowledge, just as full beliefs can, does important theoretical work. It allows us to make sense of run-of-the-mill knowledge ascriptions like (8)-(10), as well as ascriptions of scientific knowledge like (17)-(19). And it allows us to give a satisfying, unified explanation of the epistemic incorrectness in both qualitative and probabilistic Gettier cases. Moss argues that it does other important work as well, *e.g.*, it allows us to develop more plausible knowledge norms for action and decision. And we could go on. It does even more.[16] On top of that, Moss dispels some

appropriate modal condition, *e.g.*, sensitivity, or safety. For an introduction to the literature on *epistemic luck*, see Pritchard (2005, 2008).

[16]For example, probabilistic knowledge plausibly helps us sort genuine learning experiences from 'mere cognitive disturbances'. To illustrate, suppose an agent has an experience $\mathscr{E}$ that makes her certain of a proposition D (and nothing else). When does $\mathscr{E}$ count as a genuine *learning* experience, involving the acquisition of new evidence, as opposed to a pathological episode of some sort — a 'mere cognitive disturbance'? (Getting this question right is important. Different inferences are warranted in the two cases.) Fans of $E = K$ will say: exactly when $\mathscr{E}$ involves her *coming to know* D. The same question arises for non-dogmatic learning experiences. Suppose an agent has an experience $\mathscr{E}'$ that induces a Jeffrey-shift, *i.e.*, sets her credences over a partition in a particular way. Again you might ask: When does $\mathscr{E}'$ count as a genuine *learning*

prima facie serious concerns about probabilistic knowledge, *e.g.*, that accepting it forces us to reject some truisms about knowledge: that it is factive, safe, and (perhaps) sensitive. If this is all correct, it provides compelling reason to countenance probabilistic knowledge.

Our aim now is to identify necessary conditions on probabilistic knowledge (§2). Then we will develop formal tools for choosing credences that help to ensure those conditions are met, at least in an interesting range of contexts of inquiry (§3-6). If successful, these tools will yield good candidates for probabilistic knowledge. They will take us some way toward securing such knowledge. Finally, we will draw out some general lessons about the nature of probabilistic knowledge (§7).

2 Cognitive Ability

2.1 The Relationship Between Ability and Luck

Pritchard (2010, 2012) defends an anti-luck virtue epistemology. On this view, knowledge is belief that satisfies two conditions: an *anti-luck condition* and an *ability condition*.

Anti-Luck Condition. Knowledge is incompatible with luck, in the sense of being *safe*: if one knows, then one's true belief could not easily have been false. (*cf.* Pritchard (2010), p. 52)

Ability Condition. Knowledge requires cognitive ability, in the sense that if one knows, then one's cognitive success (the truth of one's belief) is the product of one's cognitive ability.[17,18] (*cf.* Pritchard (2012), p. 248)

experience, involving the acquisition of new evidence, as opposed to a 'mere cognitive disturbance'. One hypothesis: exactly when those new credences (the direct result of the Jeffrey-shift) constitute *probabilistic knowledge*.

[17]Pritchard endorses a slightly weaker ability condition, *viz.*, that one's cognitive success must be *to a significant degree creditable* to one's cognitive ability. *Cf.* (Pritchard, 2012, p. 273).

[18]Of course, one might understand the ability condition in a more or less demanding way. For example,

Knowledge-first theorists will be pessimistic about *analyzing* knowledge in this way. But they might, nevertheless, agree that the anti-luck and ability conditions specify important properties of knowledge, properties that make knowledge valuable. Similarly, we might suggest that suitably modified anti-luck and ability conditions specify important properties of *probabilistic* knowledge, even if they do not provide the building blocks of an analysis.

Probabilistic Anti-Luck Condition. Probabilistic knowledge is incompatible with luck, in the sense of being *safe*: if some of your credences constitute knowledge, then they could not easily have been wildly inaccurate.

Probabilistic Ability Condition. Probabilistic knowledge requires cognitive ability, in the sense that if some of your credences constitute knowledge, then their success (the accuracy of one's credences) is the product of cognitive ability.

Typically, these conditions run together (Sosa, 2007, pp. 28-9).[19] For example, typically if an election forecaster has a high credence that candidate A will beat candidate B, and her credence is not only accurate, but accurate because of her cognitive ability or skill (skill in assessing the current polling data, extracting the right general lessons from previous elections, etc.), then it will be safe as well. It could not easily have been inaccurate. Suppose the election is, in fact, proceeding fairly normally, or typically. There is no cockamamie plot to steal the election for B, foiled at the last minute, or anything of the sort. So her data (polling data, data about election dynamics, etc.) could not easily have been wildly

epistemic internalists might be inclined toward the view that the truth of one's belief is the product of one's cognitive ability only if one possesses reflectively accessible epistemic support for that belief. Epistemic externalists, alternatively, might insist that a wide range of reliable belief formation processes manifest cognitive ability even in the absence of reflectively accessible support. We take a broadly externalist approach in what follows. And we focus largely on one specific type of cognitive ability (*cf.* §2.2, especially fn. 22). Nevertheless, our discussion is pertinent for both internalists and externalists. It teaches us an important lesson about the nature of cognitive ability even if it does not, as the internalist maintains, tell the *whole* story about what cognitive ability consists in.

[19] See also Greco and Turri (2013), §6.

misleading. Then, in view of her skill at assessing that data, her high credence that *A* will beat *B* could only be inaccurate (*A* loses to *B*) if whole districts miraculously flip parties, or something similar. And worlds in which whole districts miraculously flip parties are distant possibilities. So her credence could not easily have been inaccurate. Indeed, the anti-luck and ability conditions seem so intimately related that some epistemologists presuppose that, once spelt out correctly, one will entail the other.[20]

But this is a mistake. As intimately related as they are, it would be wrong to suppose that one entails the other. The anti-luck and ability conditions impose logically independent epistemic demands on our full beliefs, as Pritchard emphasizes (Pritchard, 2012, p. 249). The same is true of credences. Your credences can be safe, but not accurate because skillfully produced. And they can be accurate because skillfully produced, but not safe. Imagine, for example, that the forecaster's circumstances are *not* normal. A large portion of the pro-candidate-*A* electorate could easily have had soporific drugs slipped into their coffee the morning of the election. But the goons tasked with delivering the drugs happened to slip on the ice and break their vials (or something similarly outlandish). Then our forecaster's credence is accurate because skillfully produced, but not safe. It could easily have been wildly inaccurate (because the goons could easily have not slipped).

Alternatively, imagine that our election forecaster has no clue how to assess polling data. But a crafty benefactor intent on building up her self-confidence is tracking her credence (by monitoring her predictions and so on), and paying off voters so as to *make* that credence accurate. Then it is accurate, and moreover safe, but not accurate because skillfully produced.

So neither the probabilistic anti-luck condition nor ability condition *entails* the other. Still, they hang together in a way that can and ought to guide our search for helpful formal tools, tools meant to help us secure probabilistic knowledge. In normal circumstances,

[20]See Sosa (1999). For additional discussion, see Pritchard (2012), pp. 248-9.

we can mitigate dependence on luck — make our credences safe — by reasoning skillfully. We can reason from our evidence in just the right (skillful) way, so that *its* character is paramount for explaining our success (the accuracy of our credences). And if we do, then normally we could not easily fail (arrive at highly inaccurate credences). The reason: normally our evidence could not easily be highly misleading.

The lesson: formal tools designed to yield *skillfully produced* credences will typically turn out to be doubly valuable. They will help ensure that both the ability and anti-luck conditions are satisfied. In contrast, when these conditions do *not* hang together in the normal way — when reasoning skillfully is not enough to mitigate dependence on luck — savvy prior construction will likely not help *however we proceed*. Formal tools for selecting credences simply will not, however well designed, stave off safety-undermining goons with drugs.

We will focus on the probabilistic ability condition then. Our aim: develop a new kind of objective Bayesianism designed to deliver credences that are not only reasonably likely to be accurate, in a wide range of contexts of inquiry, but whose accuracy is, to the greatest extent possible, a product of cognitive ability. In normal circumstances, these credences will be safe as well. This is the best available means, it seems, to the end of securing credences that are eligible candidates for constituting probabilistic knowledge.[21]

2.2 An Account of Cognitive Ability

What is it for the accuracy of an agent's credences to be the product of cognitive ability? One tempting proposal: it is just for them to be produced by some reliable credence-

[21]Given our focus, we will forgo further discussion of the probabilistic anti-luck condition, and any pressing concerns, *e.g.*, concerns about what could possibly ground the relevant (contextually sensitive) threshold, *i.e.*, the threshold below which the accuracy of your credences cannot drop, across some range of nearby worlds, if those credences are to amount to knowledge. *Nota bene*: If Moss' approach is correct, these concerns are moot. She explores general factivity, anti-luck (safety), and sensitivity conditions on knowledge (qualitative and probabilistic) which avoid postulating such a threshold (Moss, 2013, pp. 17-20).

forming mechanism, and for their accuracy to be explained by the exercising of that mechanism. Suppose, for example, that a doctor examines a patient, acquires a large mass of clinical data (the results of blood tests, a lumbar puncture, etc.), and uses it, together with her prior data (information about which symptoms correlate with which diseases, etc.), to arrive at a high credence that the patient has multiple sclerosis. Suppose that her high credence is rather accurate (the patient does have MS), and her reasoning process is reliable. Suppose finally that the fact that she reasoned so reliably from all of this data explains why her credence is accurate to the particular degree that it is. (It is not the case that, despite her impeccable reasoning, her credence is accurate simply by a stroke of luck.) Then the accuracy of her high credence is the product of cognitive ability, on this proposal.

Unfortunately, this proposal cannot be correct. Reasoning that manifests cognitive skill often produces equivocal, inaccurate credences. So such reasoning need not be particularly reliable. It need not tend to produce highly accurate credences. Suppose that our doctor has almost no access to the patient, and so cannot gather much clinical data. Perhaps the patient locked herself in her flat, and all the doctor can do is talk to her through the door. Our doctor might well spread her credence fairly evenly over a range of mutually exclusive hypotheses about the patient's illness in a case like this. And such credences are bound to be fairly inaccurate (after all, they are spread fairly evenly over mostly false hypotheses). Still, those credences are plausibly the product of cognitive ability. They are (more or less) the same credences that any skilled, sane doctor would have. The upshot: in a wide range of cases, skillful reasoning produces fairly equivocal, inaccurate credences. So such reasoning is not highly reliable. (It is not *anti*-reliable either, of course.) High reliability is too much to demand of skill-manifesting credence-forming mechanisms.

Instead, we should say that *having cognitive skill is a matter of reasoning in such a way that your evidence explains the accuracy (or inaccuracy) of your credences to the greatest degree possible* (even if reasoning in this way fails to give you a positively *high* chance of securing accurate credences, *e.g.*, because your evidence is too scant, unspecific, etc.).[22] The accuracy of your credences is the result of cognitive skill or ability, on this view, exactly to the extent that the reasoning process that produces them makes evidential factors (how weighty, specific, misleading, etc., your evidence is) comparatively important for explaining that accuracy, and makes non-evidential factors comparatively unimportant.[23]

This account rightly recognizes that cognitive skill is valuable *not* because it always provides you with a high chance of securing highly accurate credences (it does not, particularly when evidence is scant); rather, cognitive skill is valuable because it makes extraneous, non-evidential factors — factors that have no bearing on the character of your evidence — *irrelevant* to your chance of securing such credences. When you reason from your evidence in a way that manifests cognitive skill, your chance of success (securing accurate credences) depends not on whether some non-evidential factor turns out this way or that, but on whether your evidence is weighty or flimsy, specific or unspecific, misleading or not, etc.

In addition to respecting this insight about the value of cognitive skill, it also explains our considered judgments about whether such skill is in play in a range of cases. Suppose, for example, that a middle aged man comes into the ER with chest pain. Two doctors, Jim and Betsy, observe him. They both have the same prior evidence, let's imagine (a brief description of the man's symptoms, together with background medical knowledge). And it fails to discriminate between various competing hypotheses, *e.g.*, that the patient's

[22]For the purposes of our discussion, we will understand cognitive skill narrowly as *skill at securing accurate credences*. This is not to deny, however, that the exercising of perceptual skills, motor skills, etc., involve the manifesting of cognitive skill, understood more broadly, as they surely do.

[23]Posterior credences are a product of both an agent's prior credences and her updating rule. So arriving at posteriors *skillfully* requires having the right sort of priors/updating rule *pair*. I will assume, though, that the agents we consider update by conditionalization, for the sorts of reasons outlined in Greaves and Wallace (2006), Leitgeb and Pettigrew (2010), and Easwaran (2013) (at least when responding to traditional, dogmatic learning experiences, *i.e.*, learning with certainty). They do not update by repeated instances of MaxEnt (see §5), as proposed by Jon Williamson (2010, ch. 4), or some alternative policy. So I will suppress talk of updating rules in what follows.

chest pain is caused by a heart attack, that it is caused by hyperventilation, etc. Both Betsy and Jim's prior credences, *i.e.*, their credences prior to acquiring new clinical data, are consistent with the constraints imposed by their prior evidence, let's stipulate. But Jim's credences also reflect a hunch. He 'feels it in his bones' that the patient's chest pain is a result of hyperventilation, and so lends much more credence to that hypothesis than to the others. Betsy, on the other hand, spreads her credence much more evenly over all of the relevant hypotheses.

To be clear, Jim is *not* recognizing some ineffable signs that Betsy is missing (in the way that Dr. House might). If he were, he would have more evidence than Betsy, contra our assumption (even if he could not express that evidence to anyone). The bias in Jim's prior reflects a 'mere hunch'.

Importantly, concentrated priors like Jim's are rather *resilient* with respect to a wide range of data.[24] Learning something new does not alter them much (for many new data items). Betsy's prior, in contrast, is much more malleable, much more prone to change in the face of new data. So when Betsy runs her diagnostic tests and updates her prior, she revises her credences quite a bit. In contrast, when Jim updates on the same clinical data, he revises his credences fairly minimally.

Suppose that Betsy and Jim are both successful. Their updated credences are fairly accurate. They both end up concentrating most of their credence on the hyperventilation hypothesis, and the patient is, in fact, suffering chest pain as a result of hyperventilation. But Jim is a little closer to certain. His credences are a little more concentrated. So Jim's posterior is a little more accurate than Betsy's. Still, his accuracy seems to be less a product of cognitive skill than Betsy's, and more a product of his lucky hunch.

[24] A distribution p is resilient with respect to a datum D to the extent that the result of conditioning p on D, *i.e.*, $p_D(\cdot) = p(\cdot|D)$, is close to p. In §5.2, I introduce one particularly attractive divergence between distributions: Cramér-von Mises distance, $\mathfrak{C}$. If we use Cramér-von Mises distance as our measure of 'closeness' when explicating resilience, then we can say: a distribution p is resilient with respect to a datum D to the extent that $\mathfrak{C}(p, p_D)$ is close to zero.

The current account explains this judgment. Because Jim's prior is biased strongly in favor of the hyperventilation hypothesis (reflecting his hunch), it is fairly resilient with respect to new data. As a result, *his posterior accuracy depends a great deal on his prior accuracy*. His high degree of posterior accuracy is *explained*, in no small part, by the fact that his prior credences were quite accurate (his hunch was correct). This is constitutive, on the proposed account, of failure to manifest cognitive skill. Skilled reasoning, on this view, mitigates the explanatory relevance of non-evidential factors, such as fortuitous prior accuracy, so that they are no more relevant than required by one's prior evidence.[25]

If this is right, then we can recast the probabilistic ability condition as follows:

***Probabilistic Ability Condition*★.** Probabilistic knowledge requires cognitive ability, in the following sense: if some of your credences constitute knowledge, then your evidence explains their accuracy to the greatest degree possible.

The hope now is to find some way of sifting credal states that satisfy this ability condition from ones that do not. My plan is to identify a particular 'summary statistic' that tracks the amount of cognitive ability that a credal state manifests. I will then use this statistic to develop a novel kind of objective Bayesianism that will help us to secure credences that are not only reasonably likely to be accurate, at least in a wide range of contexts of inquiry, but whose accuracy is, to the greatest extent possible, a product of cognitive ability. Such credences are particularly good candidates for probabilistic knowledge.

3 Chance and Explanation

To identify our summary statistic, we need to say a bit more about *how* cognitive skill manages to make evidential factors (how weighty, specific, misleading, etc., your evidence

[25]Prior evidence E can require a factor F to be relevant to explaining posterior accuracy to at least degree k in the following sense: F is relevant to at least degree k for *every* prior p in the set of distributions $\mathcal{C}$ consistent with the constraints imposed by E.

is) comparatively important for explaining the accuracy of your credences, and make non-evidential factors (lucky hunches, etc.) comparatively unimportant.

Foreshadowing a bit (in an imprecise, but illustrative way), here is what we will find: Past facts help to explain future events just in case they also help to explain the *chances* of those events at intervening times (supposing those chances are defined).[26] So, if wiggling the past fact does not make the chance of the future event wiggle, so to speak, then the past fact does not *explain* the future event. This is, more or less, how cognitive skill works its magic. It makes it so that wiggling (or altering, in a Lewisian sense) certain past facts — *e.g.*, the fact that a hunch of yours turned out to be spot-on (or wildly-off) — does not wiggle (or alter) the *chance* of a certain future event — *e.g.*, arriving at accurate credences after performing an experiment, gathering some data, and updating on that data.[27] And in that case, the past fact (about prior accuracy) plausibly does not *explain* the future event (attaining posterior accuracy). The character of your data — how weighty, specific, misleading it is, etc. — explains it instead. And this is just what is required for the future event to be the product of *cognitive skill*, on our view.

To fill in the details of this story — about how cognitive skill makes evidential factors important for explaining accuracy, and makes non-evidential factors unimportant — start by considering an instructive case involving *practical* skill:

Mars Rover. The new Mars rover is beginning its descent through Mars' atmosphere. At the outset, it is blind. A bulky heat shield blocks its sensors. But before long it will eject the heat shield, release its supersonic parachute, and slow down. At that point, its sensors will make various readings, and it will maneuver its way to the

[26] More carefully, past facts help to *causally* explain future events *in the sense of figuring into the full, complete causal explanation of those future events*, just in case they also help to explain the *chances* of those events at intervening times (supposing those chances are defined). See the General Argument on p. 21 for further discussion.

[27] An alteration of one event is, to a first approximation, another event that differs slightly in when or how it occurs. See (Lewis, 2000, p. 188). Altering one event C fails to alter another E if the following is true: had any one of a range of alterations $C_1, \ldots, C_n$ of C occurred, E would have occurred just the same.

> landing site. If the rover is equipped with a guidance program that makes it *skilled at landing*, then its chance of success (touching down close to the target) would be (more or less) the same (or invariant), whether it happens to emerge from its initial (blind) descent directly above the landing site, or 1/2 mile to the north, or 3/4 miles to the northeast. Its skill will make it so that (within reasonable bounds) facts about its initial proximity to the site — facts that it has no information about during the blind part of its descent (non-evidential factors) — have little to no impact on its chance of success. As a result, the rover's initial proximity to the landing site (within reasonable bounds) will plausibly be irrelevant for explaining why that chance of success is what it is. In turn, it seems, it will also be irrelevant for explaining why the rover achieves whatever actual degree of success that it does.

Facts about the rover's initial proximity to the landing site (*i.e.*, its proximity upon emerging from initial descent) are not reflected in its prior (pre-heat-shield-ejection) evidence; they do nothing to explain the character of that evidence (the rover is blind during initial descent). Initial proximity is, in this sense, a *non-evidential factor*. The rover's skill at landing makes this sort of non-evidential factor irrelevant twice over.[28] It makes it irrelevant for explaining why the rover's *chance* of success is what it is (evidenced by the *invariance* of that chance across changes in initial proximity). It also makes it irrelevant for explaining why the rover is *actually* successful to the degree that it is (why it actually touches down close to the target or not). The two are intertwined. The rover's skill seems to make initial proximity irrelevant for explaining actual success *precisely by* making it ir-

[28]This is so whether or not conditions on Mars are kind enough for the rover to have a positively *high* chance of success (landing near the target). If the conditions on Mars are massively unpredictable, and the landing task massively complex, then our skilled rover might have a low chance of landing within, say, 0.5 miles of its landing site. (The Curiosity rover, for example, which was extremely skilled at landing, touched down 1.5 miles from its landing site.) Our rover might have a much lower chance than it would if it blindly and unskillfully rocketed due north during initial descent, burning resources before its sensors even come online (if it is in fact due south). Still, its skill at landing makes non-evidential factors irrelevant (or nearly irrelevant) for explaining why it achieves the degree of success that it does in fact achieve.

relevant for explaining the chance of success. And the reason, plausibly, that the rover's skill can operate in this way is this: No past fact (*e.g.*, about initial proximity) can help explain why the rover *actually* lands close to the target without explaining why its *chance* of landing close to the target is what it is.

This phenomenon is quite general. *Chances are explanatory foci.* Or more carefully: chances are *causal*-explanatory foci. A past fact F partially causally explains a future event E if and only if F partially explains why the chance of E takes the values that it does at intervening times (supposing the chance of E is defined).[29,30] (All the events of interest to us — touching down near a landing site, diagnosing a patient's medical condition, etc. — have causal explanations. So we can safely restrict our attention in this way.)

The General Argument that chances are causal-explanatory foci (summed up informally after the break) goes as follows.

1. *No-Skipping-Intervening-Times Thesis*. Suppose that F is a fact about the history of some (closed) causal system S at a time t^- prior to the current time t, and event E occurs in the history of S at a time t^+ after t. (So t intervenes between t^- and t^+.) Then F helps to causally explain E (in the sense of figuring into the full, complete causal explanation of E) if and only if F helps to explain one of the following:[31]

    (i) why some causally relevant part of S (*i.e.*, some part of S whose state at t

[29] A fact F about the past causally explains a future event E by explaining how or why some past event C causally influences E. Perhaps facts are *proper causal relata*, as Bennett (1988) and Mellor (1995) maintain. In that case, we might simply say: A past fact F (partially) causes a future event E if and only if F (partially) causes the chance of E to take the values that it does at intervening times (supposing the chance of E is defined). But we hope to remain neutral about the nature of causal relata. So our original formulation wins the day. Whether or not past facts causally *promote* future events, they surely causally *explain* future events, by explaining why their causal ancestors were the way they were.

[30] For our purposes, it will suffice if the possible outcomes of some interesting range of experiments are chancy, as the theoretical hypotheses we typically take to explain those outcomes posit. In that case, the events of interest for us — achieving a particular degree of (posterior) accuracy after conditionalising on the outcome of an experiment — are also chancy. And in *that* case, our summary statistic will turn out to be a good guide to the amount of cognitive ability that a credal state manifests, and useful for helping us secure probabilistic knowledge.

[31] The General Argument is modeled on (Joyce, 2007, pp. 200-202).

influences E) is in the state that it is in at t;

(ii) why the causal laws governing S (the laws that underwrite facts about causal influence in S) are what they are.

2. *Chance-Determination Thesis.* The chance of E at t is determined — and hence explained — by the states of the causally relevant parts of S at t, together with causal laws governing S. So F helps to causally explain why the chance of E is what it is at t if and only if F helps to explain either (i) or (ii).

C. *Chances as Causal-Explanatory Foci Thesis.* F helps to causally explain E if and only if F helps to explain why the chance of E is what it is at t.

Shorter: past facts F causally explain future events E if and only if they explain how or why events at intervening times influence E; but facts about such intervening events are precisely what determine the *chance* of E at intervening times; so *chances are causal-explanatory foci*; past facts F causally explain future events E if and only if they explain why the chance of E is what it is at intervening times.

The General Argument provides us with good *pro tanto* reason to think that chances are causal-explanatory foci quite generally. Of course, a full defense would come with various qualifications.[32] But we will not attempt such a defense here. It will suffice for

[32]For example, if some event C *causally inhibits* E, but E nevertheless occurs, then according to the No-Skipping-Intervening-Times Thesis, any past fact F that helps to explain why C occurred also partially causally explains E (in virtue of explaining some part of the total causal backstory of E). Consider a concrete example. Imagine that immediately prior to t, the chance of a particular particle P decaying at t is 0.01 (rather low). Nevertheless, P does in fact decay at t. Now imagine that some past fact F partially explains why P's chance of decaying is 0.01, by explaining some past event C that inhibits decay. Then according to the No-Skipping-Intervening-Times Thesis, F partially causally explains why P decays at t (since F explains C and C features in the total causal backstory of P's decaying). This, you might think, shows that the No-Skipping-Intervening-Times Thesis is false, at least without further qualification.

We might respond as follows. The practice of providing explanations involves answering contextually salient questions. What this example shows is that in *some* explanatory contexts, facts like F will not figure into a relevant answer to the why-question under discussion. But this is true of nearly any putative our purposes if the General Argument motivates the following more restricted thesis:

Restricted Chance-Explanation Thesis. In some important contexts of inquiry, both evidential factors (how weighty, specific, etc., your evidence is) and non-evidential factors (lucky hunches, etc.) help to explain why you *actually* arrive at accurate (or inaccurate) post-experiment credences if and only if they help to explain why your *chance* of arriving at such credences is what it is.

Recherché concerns about the General Argument will not call the Restricted Chance-Explanation Thesis into question. So we will assume henceforth that *chances are explanatory foci*, at least in the restricted range of contexts under consideration. Past facts (*e.g.*, about the fortuitous accuracy of your pre-experiment credences) help to explain future events (*e.g.*, settling on accurate post-experiment credences) just in case they also help to explain the chances of those events at intervening times.

This feature of chance — that it acts an explanatory focal point — might seem exotic. But it is actually quite banal. An example will help:

Asthma Drug. You have pretty bad asthma. Your doctor recommends a new drug, *BreatheEZ*. Happily, it clears right up. A few months later, you stumble across a report in the Journal of Asthma Studies. The previous clinical trials of *BreatheEZ* were flawed. In more recent trials, it failed to demonstrate a statistically significant increase in positive health outcomes compared to placebo. This, you think, is good evidence that taking *BreatheEZ* failed to affect, or explain in any way, the *chance* that

explanans. What is important is this: in *some* contexts of inquiry, facts like F *do* figure into the answer to the salient why-question. For example, if an inquisitive child asks, "Why did P decay, rather than spontaneously turning into a butterfly?" it would be perfectly appropriate to respond, "Well, because of F and a number of other factors, P had *some* chance of decaying, albeit a small one. But P had *no* chance at all of turning into a butterfly." The upshot: F is *one* of the causal-explanatory factors available to figure into answers about why P decayed, even if it in fact turns out to not be relevant in typical explanatory contexts. So it is no decisive mark against the No-Skipping-Intervening-Times Thesis that it counts F as part of the *full, complete causal explanation* of why P decayed.

> the patients in those studies had of recovering. If it had, you reason, then the new, improved clinical trials would have demonstrated a statistically significant increase in health outcomes.

Now suppose you think that you are just like those patients in all of the relevant respects, and conclude that taking *BreatheEZ* does not explain why *your* chance of recovery was what it was. Then it would be natural to infer that it fails to explain why you *actually* recovered. Maybe moving to a new city with different allergens, or something of the sort, explains your recovery. Whatever the right story is, though, *BreatheEZ* is no part of it. This inference is mediated by the chances-as-explanatory-foci thesis. No past fact (the fact that you took *BreatheEZ*) can help explain one's success (recovering from asthma) without explaining why one's *chance* of success was what it was.

We will now attempt to leverage this fact about chance — that chances are explanatory foci — to identify a 'summary statistic' of credal states that will help us sift ones that satisfy the probabilistic ability condition from ones that do not. Here is how we will proceed. Firstly, we will look for a statistic that helps us sort out, for any given credal state, what explains why its *chance* of success (posterior accuracy) is what it is, in any experimental context. Does it reflect some sort of hunch that goes beyond the available prior evidence? A hunch which explains why it has a particularly high (or low) chance of attaining a particularly high (or low) degree of posterior (post-experiment) accuracy? Or are those chance-facts explained, rather, by the character of the available experimental evidence: how weighty it is, how specific it is, etc.?

If we find such a statistic, we will be off to the races. It will help us sort out which factors explain why any prior (pre-experiment) credal state has the *chance* that it does of attaining any particular degree of posterior (post-experiment) accuracy. And because chances are explanatory foci, this will enable us to sort out which factors explain why that credal state *actually* attains the degree of posterior accuracy that it does. And this is exactly what determines whether such accuracy is the product of cognitive ability. So we will be able to sift credal states that satisfy the probabilistic ability condition from ones that do not.

4 Cognitive Ability and Objective Expected Accuracy

4.1 An Example: Scientific Inquiry

To have a concrete case in front of us, while searching for our ability-tracking summary statistic, imagine the following. A microbiologist designs and performs an experiment to adjudicate between competing theoretical hypotheses $H_1, \ldots, H_n$, *e.g.*, hypotheses about whether and how over-expression of a certain gene causes chromosomal instability in breast tumors.[33] She has prior (pre-experiment) credences for $H_1, \ldots, H_n$, which reflect both her prior evidence — information about past patients, about chromosomal instability in other sorts of tumors, and so on — as well her personal inductive hunches and quirks. And she will soon acquire new experimental data, which she will use to update those prior credences, to arrive at new, better-informed posterior (post-experiment) credences.

For precision, assume that our agent has opinions about propositions in an atomic Boolean algebra $\mathcal{F}$, which means that (i) $\mathcal{F}$ is closed under negation and disjunction, and (ii) every proposition in $\mathcal{F}$ can be expressed as a disjunction of atoms (the logically strongest elements of $\mathcal{F}$).[34,35] We can think of each atom w of $\mathcal{F}$ as the 'possible world'

[33] *Gene expression* is the transcription and translation process that turns genetic information into protein. A gene is *over-expressed* when that process goes awry, yielding a surplus of protein.

[34] The atoms w of $\mathcal{F}$ are the logically strongest elements of $\mathcal{F}$ in the following sense: for every $X \in \mathcal{F}$, w either entails X or $\neg X$, and X entails w only if $w = X$ or $X = Y \& \neg Y$ for some $Y \in \mathcal{F}$.

[35] $\mathcal{F}$ need not be finite. In some examples, though, we will assume that it is, simply for expositional ease.

in which all of the propositions X in $\mathcal{F}$ that w entails are true, and all other propositions Y in $\mathcal{F}$ are false.[36] Let $\mathcal{W}$ be the set of all such atoms or 'worlds'.[37] Let $c : \mathcal{F} \to [0,1]$ be the real-valued credence function (or credal distribution) that measures how confident our microbiologist is in propositions in $\mathcal{F}$. And let $ch : \mathcal{F} \to [0,1]$ be the true chance function, *i.e.*, the real-valued function that specifies the true chances of propositions in $\mathcal{F}$ (immediately prior to conducting the experiment).[38]

Assume also that our agent satisfies some fairly uncontroversial epistemic norms. Her credences are probabilistically coherent.[39] She updates by conditionalization, so that, when she acquires a new datum D (and nothing more), she adopts a posterior credence function $c' : \mathcal{F} \to [0,1]$ which satisfies $c'(X) = c(X|D)$ for all $X \in \mathcal{F}$.[40] And she obeys the Principal Principle. So she treats chance as an *epistemic expert*. When she learns that chance's probability for X at t is x, she straightaway adopts x as her new credence for X.[41]

[36]Thinking of atoms w of $\mathcal{F}$ as possible worlds in this way is metaphysically innocuous. The point is simply that such w *function as* doxastically possible worlds *for the agent*. They are the finest-grained descriptions of the way the world could be from her perspective.

[37]If you prefer to model propositions as sets of worlds, let $\mathcal{W}$ be the set of all possible worlds, and let $\mathcal{F}$ be a sigma-algebra of subsets of $\mathcal{W}$ (closed under complement and union).

[38]On many accounts of chance, *e.g.*, the propensity theories of Fetzer (1982, 1983) and Gillies (2000), the true chance function is only defined on a rather restricted set of propositions. On other accounts, *e.g.*, Hitchcock (2012) and (citation omitted for anonymity), even supposing the chances are well-defined, they need not be precise. Chances are often *imprecise* (modelable by a *set* of probability functions). We restrict our attention, however, to algebras $\mathcal{F}$ for which the true chances are both defined and precise.

[39]See Joyce (1998, 2009), Predd et al. (2009), and Schervish et al. (2009) for an accuracy-dominance argument that rational credence functions must be probabilistically coherent, *i.e.*, must satisfy the laws of finitely additive probability:

1. $c(\top) = 1$
2. $c(\bot) \leq c(X)$
3. $c(X_1 \vee ... \vee X_n) = c(X_1) + ... + c(X_n)$ for any pairwise incompatible propositions $X_1, ..., X_n$

Axiom 1 says that you must invest full confidence in tautologies. Axiom 2 says that you must invest at least as much confidence in any proposition X as you do a contradiction. Axiom 3 says that the amount of confidence that you invest in a disjunction $X_1 \vee ... \vee X_n$ of pairwise incompatible propositions $X_1, ..., X_n$ must be the sum of your degrees of confidence in each of the X_i.

[40]When $c(D) > 0$, $c(X|D)$ is just $c(X \& D)/c(D)$. For an expected accuracy argument for conditionalization, see Greaves and Wallace (2006), Leitgeb and Pettigrew (2010), and Easwaran (2013).

[41]For ease of exposition, we focus on a version of the Principal Principle that Pettigrew (2015, p. 137) calls the *extended temporal principle*:

Finally, make a couple of additional assumptions about the propositions in our microbiologist's algebra. She has credences about the competing theoretical hypotheses $H_1, ..., H_n$ that she hopes to adjudicate between, which we suppose are pairwise incompatible and jointly exhaustive. She has credences about the different possible experimental outcomes she might observe, or the new data she might receive, $D_1, ..., D_m$ (also pairwise incompatible and jointly exhaustive). And these two sets of propositions are related. Each hypothesis H_i specifies chances for the possible data items $D_1, ..., D_m$. To have a way of talking about them, let ch_i be the function that would specify the chances if H_i were true. So the chance that our microbiologist's experiment would yield outcomes $D_1, ..., D_m$ if H_i were true is $ch_i(D_1), ..., ch_i(D_m)$, respectively. (This notation will be helpful shortly.)

Our question is this: What can we say about the extent to which the character of our microbiologist's evidence explains the accuracy of her posterior credal state (and hence the extent to which her credences satisfy the probabilistic ability condition)? Do other factors play a significant explanatory role? In particular, do her prior credences reflect some sort of hunch that goes beyond her prior evidence (as Dr. Jim's did), so that her posterior accuracy depends, in large part, on fortuitous prior accuracy?

First, we will outline the answer. Then we will unpack it. The answer, briefly, is this: If her *objective expected posterior accuracy* would be the same (invariant), regardless of whether her *prior* credences reflect a particularly accurate hunch or not, then that hunch plausibly plays no role in explaining why she has a particularly high (or low) chance of attaining a particularly high (or low) degree of *posterior* accuracy. And if the accuracy

---

ETP. If an agent has a credence function $c : \mathcal{F} \to [0,1]$, then rationality requires that

$$c(X|T_{ch}) = ch(X)$$

for all propositions X in $\mathcal{F}$, and all possible current chance functions ch such that $c(T_{ch}) > 0$, where T_{ch} is the proposition that the current chance function is ch.

See Pettigrew (2015) for a chance-dominance argument for ETP, as well as discussion of alternative deference to chance principles.

(or inaccuracy) of her hunch does not explain why her *chances* are what they are, then it simply cannot explain why she attains the *actual* degree of posterior accuracy that she does. Chances, after all, are explanatory foci.

The upshot: if her objective expected posterior accuracy would be the same, whether her *prior* credences reflect a particularly accurate hunch or not, then the accuracy of that hunch is plausibly irrelevant for explaining why she attains whatever actual degree of *posterior* accuracy that she does. Rather, her success or failure on that front is explained primarily by the character of her evidence: how weighty, specific, misleading it is, etc. And this is just what is required for the accuracy of her posterior credences to be the product of cognitive ability.

To unpack this answer, make two observations.

4.2 Observation 1: Accuracy vs. Expected Accuracy

Firstly, we can distinguish between the *actual* accuracy of our microbiologist's prior and posterior credences — roughly, how close they are to the actual truth-values — and the *objective expected posterior accuracy* of her credences. A bit of background will help illuminate what objective expected accuracy is and why it is important. Expectations are estimates.[42] More carefully, if $\mathcal{V}$ is some (real-valued) variable — *e.g.*, the annual rainfall in New York, the amount of money in your bank account, the degree of inaccuracy of a microbiologist's credences — then $Exp_p(\mathcal{V}) = \sum_{w \in \mathcal{W}} p(w) \cdot \mathcal{V}(w)$ is the best estimate of $\mathcal{V}$, according to the probability function p, where $\mathcal{V}(w)$ is the value that $\mathcal{V}$ takes in world w (the annual rainfall in New York in w, etc.).[43] Objective expectations are the estimates determined by objective probability functions — the true *chance* function, in particular. So, for example, if *ch* is the true chance function, then its best estimate of the annual rainfall

[42] See for example Jeffrey (2004), chapter 4.

[43] Variables $\mathcal{V} : \mathcal{W} \to \mathbb{R}$ are functions from worlds to real numbers which model (measurable) quantities of interest, *e.g.*, the annual rainfall in New York (in inches), or the length of Russell's yacht (in feet).

$\mathcal{R}$ in New York is just the objective expected value of $\mathcal{R}$, $Exp_{ch}(\mathcal{R}) = \sum_{w \in \mathcal{W}} ch(w) \cdot \mathcal{R}(w)$.

Similarly, chance's best estimate of how successful our microbiologist will be — how accurate her credences will be after performing the experiment, observing the outcome, and updating on her new data — is just her *objective expected posterior accuracy*. Imagine, for example, that H_i is the true theoretical hypothesis. So ch_i is the true chance function (just prior to performing the experiment). And the chance of the experiment producing outcomes $D_1, \ldots, D_m$ is $ch_i(D_1), \ldots, ch_i(D_m)$, respectively. Now note that updating (conditioning) her prior credence function c on these bits of possible new evidence yields different posterior credence functions, $c_{D_1}, \ldots, c_{D_m}$. And each of these credence functions is (or at least could be) accurate to a different degree. So not only does chance have views about which bit of evidence our microbiologist will receive — given by $ch_i(D_1), \ldots, ch_i(D_m)$ — it also has views about how likely she is to attain various different degrees of posterior accuracy; views which are summed up by chance's best estimate of her success, *i.e.*, her objective expected posterior accuracy.

To keep track of these degrees of accuracy, and to say explicitly what our microbiologist's objective expected posterior accuracy is, let's introduce some notation. Firstly, at any world w, let "c'" denote the result of updating (conditioning) our microbiologist's prior credence function c on the experimental data that she receives in w. For example, if she receives data D_j in w (*i.e.*, w implies D_j), then "c'" denotes c_{D_j} at w. (So "c'" is a non-rigid designator.) Next, let $\mathcal{I}(c', w)$ be a non-negative real number which (in some as of yet unspecified way) measures the inaccuracy of c' at w. So if our microbiologist receives data D_j in w, then $\mathcal{I}(c', w) = \mathcal{I}(c_{D_j}, w)$.

Interpret $\mathcal{I}$ as follows. If $\mathcal{I}(c', w) = 0$, then c' is minimally inaccurate at w. Inaccuracy increases as $\mathcal{I}(c', w)$ grows larger. Finally, let $\mathcal{I}_{c'} : \mathcal{W} \rightarrow \mathbb{R}$ be the function which maps any world w to c''s measure of inaccuracy at that world, $\mathcal{I}(c', w)$. (See §5.2 for extended discussion of inaccuracy measures.)

Now we can specify the objective expected posterior accuracy of our microbiologist's credences: $Exp_{ch_i}(\mathcal{I}_{c'}) = \sum_{D_j \in \{D_1,\ldots,D_m\}} ch_i(D_j) \cdot \sum_{w \in D_j} ch_i(w|D_j) \cdot \mathcal{I}(c_{D_j}, w)$. This is ch_i's best estimate of how successful she will be. It summarizes all of the information that ch_i sees as relevant to what degree of posterior accuracy she will achieve. The closer $Exp_{ch_i}(\mathcal{I}_{c'})$ is to zero, the more likely she is (all things considered) to be successful, according to ch_i (*i.e.*, to end up with accurate post-experiment credences).

4.3 Observation 2: Counterfactual Independence as a Guide to Explanatory Irrelevance

Second observation: Chance's best estimate of how successful our microbiologist will be — or rather, how that estimate behaves across a range of counterfactual scenarios — is a good guide to what is relevant (and not relevant) for explaining why her chance of success is what it is. Imagine, for example, that she could perform her experiment using instrument A, B or C. If chance's best estimate of how successful she will be — her objective expected posterior accuracy — would be the same whether she opted for A, B or C, then which instrument she uses plausibly plays no role in explaining why that expectation is what it is. Counterfactual independence provides good (but defeasible) evidence of explanatory irrelevance.[44]

Moreover, *expectations* are typically good proxies for *whole distributions* (at least for the purposes of sorting out what does and does not help to explain one's chance of success). In many experimental contexts, if one's *expected* accuracy would stay the same, were the value of some variable to change (*e.g.*, which instrument our microbiologist uses), then

[44]Counterfactual independence only provides *defeasible* evidence of explanatory irrelevance, however. Imagine, for example, that our microbiologist uses high quality growth media A for her cell culture. If she had used lower quality growth media B or C, however, her lab-mate would have added a supplement that made it functionally equivalent to A. In that case, using A helps to (causally) explain why her expected accuracy is what it is, even though that expectation would be the same whether she opted for A, B or C. We ignore cases cases of this sort, involving preemption or trumping, in what follows. Such cases are outside the domain of applicability of the tools developed here.

one's *entire distribution* of chances over accuracy hypotheses (hypotheses about what degree of posterior accuracy one achieves) would stay the same.[45] So we can say: which instrument she uses — A, B or C — plausibly plays no role in explaining why her chance of attaining any particular degree of posterior accuracy is what it is. (We could, if we were so inclined, avoid using expectations as proxies. But doing so would be messy, and provide no additional insight. So the game is not worth the candle.[46])

More generally, if chance's best estimate of your success — your objective expected posterior accuracy — would be the same, regardless of what value some variable $\mathcal{V}$ takes, then $\mathcal{V}$ plausibly plays no role in explaining why your chance of attaining any particular degree of posterior accuracy is what it is.

Now recall our initial question. We wanted to know whether our microbiologist's prior credences reflect some sort of hunch that goes beyond her prior evidence; a hunch which explains why she has a particularly high (or low) chance of attaining a particularly high (or low) degree of posterior accuracy; a hunch which, as a result, helps explain why she attains whatever actual degree of posterior accuracy that she does. We now have the tools to sort this out.

[45]Letting *expected* accuracy go proxy for the *entire distribution* of chances over accuracy hypotheses is often harmless. Specifically, when your prior credences over candidate chance functions takes the form of a beta distribution (a very flexible class of prior distributions; see *f.n.* 53), and those chance functions themselves are binomial distributions (which will be the case, *e.g.*, when the trials of your experiment are identical and independent, and yield a sequence of 'successes' and 'failures'), then the distribution over accuracy hypotheses determined by any such chance function can be approximated by an exponential distribution. And it is easy to show that the Cramér-von Mises distance (to take one example) between any two exponential distributions is bounded by the difference between their means, or expected values. So objective *expected* posterior accuracy is invariant across a range of values for some variable only if the *entire distribution* of chances over posterior accuracy hypotheses is invariant.

[46]We could, in principle, directly examine how much one's chance distribution over accuracy hypotheses would vary across a range of values for some variable of interest (using an appropriate metric on the space of probability distributions, such as Cramér-von Mises distance; *cf.* §6). This would allow us to avoid the detour through objective *expected* accuracy altogether.

4.4 Explaining Posterior Accuracy

Suppose our microbiologist's objective expected posterior accuracy would be the same, regardless of whether her prior credences reflect a particularly accurate hunch or not; regardless whether the variable of interest for us — prior accuracy — takes a particularly high or low value. Put differently, suppose it would be the same, regardless of whether the *true* hypothesis turns out to be H_i — one that her prior credence function c initially centers a great deal of probability mass around (in which case c has a high degree of prior accuracy) — or whether it turns out to be H_j — one that c centers little mass around (in which case c has a lower degree of prior accuracy).

More succinctly, suppose:

Invariance. $Exp_{ch_i}(\mathcal{I}_{c'}) \approx Exp_{ch_j}(\mathcal{I}_{c'})$ for all H_i and H_j with $1 \leq i, j \leq n$.

Then we can say: whether c reflects a particularly accurate hunch or not (whether c attains a particularly high or low degree of prior accuracy) is irrelevant for explaining why its *chance* of attaining any particular degree of posterior accuracy is what it is.

Finally, recall that chances are explanatory foci. No past fact (*e.g.*, facts about prior accuracy) can explain a future event (attaining a particular degree of posterior accuracy) without (partially) explaining why that event had the chance of occurring that it did. So, since fortuitous *prior* accuracy is irrelevant for explaining why c's chance of *posterior* accuracy is what it is, it is also irrelevant for explaining why c attains whatever *actual* degree of posterior accuracy that it does.

Our microbiologist's posterior credences, then, are more like Dr. Betsy's than Dr. Jim's. Their accuracy is more the product of cognitive skill than lucky hunch. The reason: her *prior* credences have a special property. Their objective expected posterior accuracy is invariant across theoretical hypotheses. This allows us to tell a principled story about what does and does not explain her success (or failure) at the end of the day; her high (or low) degree of *posterior* accuracy. Her posterior accuracy (or inaccuracy) is *not* explained by fortuitous prior accuracy (or lack thereof). It is *not* explained by the accuracy (or inaccuracy) of a spot-on (or wildly-off) prior hunch. The reason: No such hunch helps to explain why her *chance* of attaining a particularly high (or low) degree of posterior accuracy is what it is. (If it did, then wiggling the accuracy of that hunch by wiggling which theoretical hypothesis is true would also wiggle her objective expected posterior accuracy.) And chances are explanatory foci. So no such hunch helps to explain why she *actually* attains a particularly high (or low) degree of posterior accuracy. Rather, her success (or failure) on this front is explained primarily by the *character of her evidence*.

This answer to our initial question has various moving parts. But once we get a grip on them, the point is simple. Consider an analogy. If chance's best estimate of how many games Ghana will win in the World Cup — Ghana's objective expected win total — would be the same, regardless of what color socks they decide to wear, then sock color plausibly plays no role in explaining why their *chance* of winning 1, or 2, or 3 games, etc., is what it is. And chances are explanatory foci. So sock color also plays no role in explaining why they *actually* win the number of games that they do.

Conversely, if Ghana's objective expected win total *would* differ depending, *e.g.*, on how many days of rest the team receives in the weeks leading up to the World Cup, then rest time plausibly *does* play a role in explaining why their chance of winning 1, or 2, or 3 games, etc., is what it is. And again chances are explanatory foci. So rest time also plays some role in explaining why they *actually* win the number of games that they do.

Similarly, if chance's best estimate of how successful our microbiologist's credences will be — her objective expected posterior accuracy — would be the same, regardless of how accurate a hunch those credences happen to encode, then that hunch plausibly plays no role in explaining why her *chance* of being accurate to degree x, y, or z is what it is. And chances are explanatory foci. So having a spot-on (or wildly-off) hunch also plays no role in explaining why she *actually* attains the degree of posterior accuracy that she does. Our microbiologist's actual degree of posterior accuracy is, rather, explained primarily by the character of her evidence: how weighty it is, how specific it is, whether or not it is misleading, etc. *And this is just what is required for the accuracy of her posterior credences to be the product of cognitive ability.*

We now have the 'summary statistic' of credal states that we have been looking for; the statistic that will help us sift credal states that satisfy the probabilistic ability condition from ones that do not. It is: *variance in objective expected posterior accuracy across theoretical hypotheses*. (We will say more about how to measure variance in §6.) The smaller this statistic, the greater the extent to which posterior accuracy is the product of cognitive ability.

My aim now is to use this statistic to help us secure nice posterior credences — credences that are not only reasonably likely to be accurate, but whose accuracy is the product of cognitive ability — in a wide range of inference problems. *Inference problems*, of the sort familiar from classical statistics, arise in structured, but nonetheless fairly ubiquitous contexts of inquiry (in the sciences, in engineering, etc.): we have competing hypotheses we wish to adjudicate between; we will soon acquire new (experimental) data that we can use (together with our prior data) to help us so adjudicate; and we know enough about the data-generating process (the experiment) to know which data items could be produced, and what the chances of receiving each of the various possible data items would be, were these different theoretical hypotheses true (typically as a result of careful experimental design). The formal tools developed here will take us some way toward securing probabilistic knowledge in many of these important contexts of inquiry.

It is worth noting, before we move on, that the central examples in §5-6 will involve a particularly simple type of inference problem: flipping a coin in order to adjudicate between competing hypotheses about its bias. But one might wonder: What does this tell us about the more complicated inference problems that arise in the sciences, etc.? The answer: a surprising amount. To the extent that the formal tools developed here help secure probabilistic knowledge in simple, coin-flipping problems, they will also help us secure such knowledge in *any* inference problem that involves performing a fixed number of identical and independent trials of an experiment with only finitely many possible outcomes. Quite a lot of inference problems in the sciences, engineering, etc., are like this (though, of course, many are not). See §10 for further discussion.

5 The Maximum Entropy Method

5.1 A Case Study

Consider an agent with opinions about propositions in an atomic Boolean algebra $\mathcal{F}$. She is planning to perform an experiment aimed at adjudicating between competing theoretical hypotheses, $H_1, \ldots, H_n$ (pairwise incompatible and jointly exhaustive). And her experiment might yield data items, $D_1, \ldots, D_m$ (also pairwise incompatible and jointly exhaustive). Suppose that $\mathcal{F}$ contains $H_1, \ldots, H_n$ and $D_1, \ldots, D_m$.

According to one popular brand of objective Bayesianism — the *maximum entropy principle*, or *MaxEnt* — our agent should take her prior information into account in inquiry as follows:

- Firstly, she should summarize that prior information by a system of constraints on prior credence functions $b : \mathcal{F} \to [0, 1]$, *e.g.*:
  - $b(H_i) \geq b(H_j)$ (*i.e.*, H_i is at least as probable as H_j)
  - $0.6 \leq b(H_i) \leq 0.9$ (*i.e.*, H_i is probable to at least degree 0.6 and at most 0.9)
  - $b(H_i|D) \geq b(H_j|D')$ (*i.e.*, H_i conditional on D is at least as probable as H_j conditional on D')

which we represent by a set $\mathcal{C}$ of credence functions, *viz.*, the set of all minimally rational credence functions, b (*i.e.*, b that satisfy Probabilism, the Principal Principle, etc.), consistent with those constraints.

- Secondly, she should adopt some prior credence function $c : \mathcal{F} \to [0,1]$ or other that maximizes Shannon entropy over $\mathcal{C}$.[47] (Under a broad range of conditions, this 'MaxEnt prior' will turn out to be unique.[48])

If she has no prior information about $H_1, ..., H_n$ at all — her evidence fails to constrain her prior credences altogether — then the maximum entropy distribution is just the *uniform* distribution.[49] So MaxEnt agrees with Laplace's Principle of Indifference:

[47] If the credence functions c in $\mathcal{C}$ are defined over only finitely many theoretical hypotheses, then their 'entropy' or 'uninformativeness', on the MaxEnt approach, is measured by their *Shannon entropy*,

$$H(c) = -\sum_{1 \leq i \leq n} c(H_i) \cdot log(c(H_i))$$

If they are defined over uncountably many theoretical hypotheses, parameterized by $\theta_1, ..., \theta_n$, then it is standardly measured by their *differential entropy*,

$$h(c) = -\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, ..., x_n) \cdot log(f(x_1, ..., x_n)) dx_n \ldots dx_1$$

where f is c's joint probability density function for $\theta_1, ..., \theta_n$. See footnote 52 for more information about probability density functions.

[48] Since Shannon entropy H is a continuous, strictly convex, real-valued function, there will be a unique MaxEnt prior whenever $\mathcal{C}$ is closed (*i.e.*, contains all of its limit points), bounded (*i.e.*, is a subset of a ball of finite radius), and convex (*i.e.*, is closed under mixtures). This will be the case, for example, when the agent's prior evidential constraints specify a (closed) range of expected values for some number of variables (the expected annual rainfall in New York is between 30 and 60 inches (inclusive); the expected price of Tesla's stock at the end of the quarter is between 210 and 230 (inclusive); etc.). When there is no unique MaxEnt prior — either because (i) many priors maximize entropy, or (ii) none do — objective Bayesians such as Jon Williamson (2010) recommend either (iii) adopting some prior or other that maximizes entropy (in the former case), or (iv) adopting some prior or other with "sufficiently high" entropy (in the latter case), where what counts as sufficiently high depends on pragmatic considerations.

[49] More carefully, when our evidence provides no constraints on prior credences $c : \mathcal{F} \to [0,1]$, and *there exists* a countably additive uniform distribution on $\mathcal{F}$, *then* the MaxEnt distribution is just the uniform distribution. In certain contexts, however, the MaxEnt prior exists while a countably additive uniform prior does not. It is well known, for example, that if $\mathcal{F}$ is an infinite-dimensional space, then there is no Lebesgue measure on $\mathcal{F}$, and hence, no analogue of the standard uniform distribution on $\mathcal{F}$. But there is often a MaxEnt prior on such $\mathcal{F}$. See, for example, Furrer et al. (2011).

POI. If an agent has no information about hypotheses $H_1, ..., H_n$, and so no reason to think that any one is more or less probable than any other, then she ought to be equally confident in each: $c(H_i) = c(H_j)$ for all H_i and H_j.

Indeed, we can think of MaxEnt generally as making the same sort of recommendation as POI. POI says that in one special case — when your evidence imposes no constraints on your credences — those credences should be uniform. MaxEnt says that in *all* cases your credences should be as *close* to uniform as possible, while satisfying the constraints imposed by your prior evidence.[50]

Jaynes offered an information-theoretic rationale for MaxEnt. Shannon entropy, Jaynes thought, is uniquely reasonable as a measure of the informativeness of a prior. So "the maximum-entropy distribution may be asserted for the positive reason that it is... maximally noncommittal with regard to missing information," he concluded (Jaynes, 1957, p. 623). But many find this argument uncompelling.[51] We will explore, briefly, whether MaxEnt has an alternative rationale — whether it delivers credences that are eligible candidates for constituting probabilistic knowledge.

Imagine that a bookie hands you and your friend a coin, and offers you a bet: win \$$w$ if it comes up heads; lose \$$l$ if not. Neither of you have any prior evidence about the coin's bias, *i.e.*, the chance that it will come up heads. But you are allowed to flip the coin for a bit — 14 times, perhaps — in order to gather some new evidence before deciding whether or not to take the bet. You adopt the maximum entropy prior, which given your dearth of evidence is just the uniform distribution u. So you spread your credence evenly over the various competing hypotheses about the bias of the coin — hypotheses of the

[50]If c uniquely maximizes Shannon entropy over $\mathcal{C}$, then it is as close as possible to the uniform distribution u in the following sense: for any other $b \in \mathcal{C}$, the *Kullback-Leibler divergence* of u from b, $\mathcal{D}_{KL}(b, u) = \sum_w b(w) \cdot log(b(w)/u(w))$, is greater than the KL divergence of u from c.

[51]See for example Seidenfeld (1986) and Joyce (2009, pp. 284-5).

form $B = x$ (read: the coin's bias B is x).[52]

Your friend, in contrast, adopts a more concentrated prior credence function b (*e.g.*, a *beta distribution* which centres most of its probability density around the hypothesis the coin's bias is 5/7; see figure 1).[53] Your friend's prior credences, as a result (like Dr. Jim's in §2), are rather resilient with respect to a wide range of data. Flipping the coin a few times and observing its outcome does not alter them much.[54] In contrast, *your* prior credences (like Dr. Betsy's) are much more malleable, much more prone to change in the face of new data. When *you* flip the coin and observe its outcome, you revise your credences quite a bit.

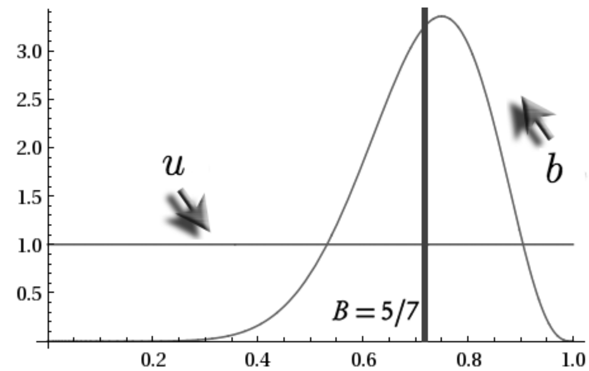

Figure 1: The uniform distribution u and more concentrated beta distribution b (with shape parameters $\alpha = 10$ and $\beta = 4$) over hypotheses of the form $B = x$.

The upshot: your friend's success — whatever degree of posterior accuracy she actually ends up with after flipping the coin and observing its outcome — seems to be more a

[52]More carefully, your credences about the coin's bias B are given by $u(s \leq B \leq t) = \int_s^t f(x)dx$ for $0 \leq s < t \leq 1$, where f is the uniform *probability density function* f, *i.e.*, the function that assigns the same probability density to each hypothesis, $B = x$ ($0 \leq x \leq 1$), *viz.*, $f(x) = 1$.

A distribution u's density function f, intuitively, centers more probability density on hypotheses that u sees as more plausible. The probability that u assigns to the *true* hypothesis being in some set S is equal to the volume under f on that region S (which is higher the more mass f attaches to hypotheses in S).

[53]Beta distributions b are parameterized by two quantities, α and β. These 'shape parameters' determine which hypotheses the distribution b focuses its probability mass on. The 'concentration parameter', $\alpha + \beta$, corresponds roughly to how 'peaked' b is around its mean. The larger (smaller) α is compared to β, the more b focuses probability mass on chance hypotheses $B = x$ with $x \approx 1$ ($x \approx 0$). Beta distributions are fairly computationally tractable, and form a very flexible class of distributions. In fact, *any* prior can be approximated by a mixture of beta distributions (Walley, 1996, p. 9). And beta distributions have nice dynamic properties as well. For example, they continue to be beta distributions when updated on various sorts of data (they are *conjugate priors* to the binomial, negative binomial and geometric likelihood functions). For these reasons, we will focus on them in many of our examples.

[54]See footnote 62 for an illustration of b's resilience.

product of her lucky (or unlucky) hunch (*viz.*, that the coin's bias is approximately 5/7) than yours. It is less a product of cognitive skill.

Our summary statistic bears this out, as we will demonstrate shortly. Here is what we will show. Chance's best estimate of how successful u will be — u's objective expected posterior accuracy — varies fairly minimally across hypotheses about the coin's bias: much less than b's. It would take nearly the same value, regardless of how accurate a hunch u happens to encode (regardless of whether the *true* hypothesis ends up being one that makes u have a relatively high degree of prior accuracy or not). So having a spot-on (or wildly-off) hunch about the coin's bias plays less of a role in explaining why *your* chance of securing particularly accurate (or inaccurate) posterior credences is what it is. Consequently, it plays less of a role in explaining why you *actually* end up with particularly accurate (or inaccurate) posterior credences. The accuracy of *your* posterior credences is explained, to a much greater degree, by the character of your evidence: how many times you flip the coin (how weighty your evidence is), whether the frequency of heads in your sequence of coin flips ends up being close to the coin's true bias (how misleading your evidence is), etc. And this is just what is required for the accuracy of your posterior credences to be more a product of cognitive skill than your friend's.

For concreteness, let's measure (in)accuracy by an *epistemic scoring rule* or *inaccuracy score*. An inaccuracy score is a function $\mathcal{I}$, which maps credence functions c and worlds w to non-negative real numbers, $\mathcal{I}(c, w)$. $\mathcal{I}(c, w)$ measures how inaccurate c is if w is actual. If $\mathcal{I}(c, w)$ equals zero, then c is minimally inaccurate (maximally close to the truth) at w. Inaccuracy increases as $\mathcal{I}(c, w)$ grows larger.

Reasonable inaccuracy scores satisfy a range of constraints (see Joyce (1998, 2009) and Pettigrew (2015)). But instead of detailing these constraints, we will simply focus our attention on a particularly attractive inaccuracy measure: *the continuous ranked probability score* (CRPS). Readers that are uninterested in the details of the CRPS (which are fairly formal), or the reasons for focusing on it, can skip ahead to §5.3.

5.2 Continuous Ranked Probability Score

The continuous ranked probability score (CRPS), as detailed in, *e.g.*, Hersbach (2000) and Gneiting and Raftery (2007), provides one way of measuring the inaccuracy of *continuous credal distributions*. Your credal distribution is *continuous* when you have credences about variables (*e.g.*, the annual rainfall in New York) which can take uncountably many values (rather than credences about, *e.g.*, a finite number of propositions, which take only two values: 1 at worlds where they are true, and 0 where false).[55] So, in particular, if you have credences about the bias of a coin (a variable which can take any of the uncountably many real values between 0 and 1) — as you do in the case at hand — then your credal distribution is continuous.

According to the CRPS, the inaccuracy at a world w of a continuous credal distribution p over hypotheses about the value of a variable $\mathcal{V}$ is given by:

$$\textbf{(CRPS)}\quad \mathcal{I}(p,w) = \int_{-\infty}^{\infty} |P(x) - \mathbb{1}(x \geq \mathcal{V}(w))|^2 dx$$

where $\mathcal{V}(w)$ is the value of $\mathcal{V}$ at w, and $\mathbb{1}$ is the indicator function, defined as follows:[56]

$$\mathbb{1}(x \geq \mathcal{V}(w)) = \begin{cases} 0 \text{ if } x < \mathcal{V}(w) \\ \\ 1 \text{ if } x \geq \mathcal{V}(w) \end{cases}$$

[55] Strictly speaking, a distribution $p : \mathcal{F} \to [0,1]$ over hypotheses about the values of variables $\mathcal{V}_1, ..., \mathcal{V}_n$ is called continuous when its cumulative distribution function $P : \mathcal{F} \to [0,1]$ is continuous. The CDF P of p is defined by $P(x) = p(\mathcal{V} \leq x)$, and specifies the probability (according to p) that $\mathcal{V}$ takes a value less than or equal to x. (The notion of continuity in play here is the standard topological notion. The relevant topology on $\mathcal{F}$ is just the topology generated by modeling the atoms of $\mathcal{F}$, which take the form $\mathcal{V}_1 = x_1 \& ... \& \mathcal{V}_n = x_n$, as points $\langle x_1, ..., x_n \rangle$ in $\mathbb{R}^n$, and lifting the standard open ball topology on Euclidean n-space.) Some authors reserve the term for distributions with absolutely continuous CDFs.

[56] When we say that $\mathcal{V}(w) = x$ is the value of $\mathcal{V}$ at w, this is shorthand for the following: the atom w entails the truth of $\mathcal{V} = x$ and falsehood of $\mathcal{V} = y$ for all $y \neq x$.

The remainder of §5.2 is devoted to unpacking this definition, and to exploring the various desirable properties of the CRPS. But firstly, a bit of background.

The CRPS falls out of a very general framework for thinking about accuracy. Pettigrew (2013b) introduces the framework as follows:

> The idea is that the epistemic utility (accuracy) of a credence function at a world ought to be given by its 'proximity' to the credence function that is 'perfect' or 'vindicated' at that world — that is, the credence function that perfectly matches whatever it is that credence functions strive to match. Thus, we need to identify, for each world, the 'perfect' or 'vindicated' credence function; and we need to identify the correct measure of distance from each of these credence functions to another credence function. (Pettigrew 2013b, p. 899; parenthetical mine)

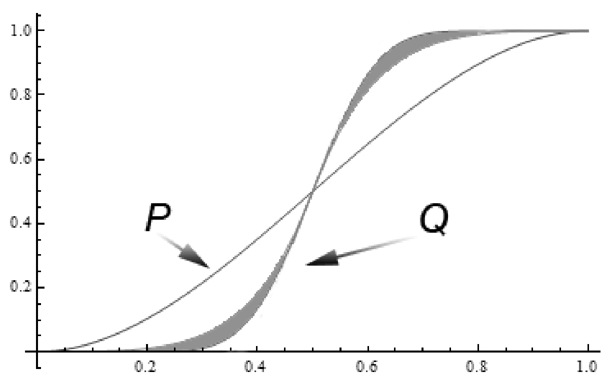

Figure 2: The Cramér-von Mises distance between beta distributions, p and q, is given by the average squared difference between their CDFs, P and Q (shaded area).

The CRPS takes the perfect or vindicated credence function at w, υ_w, to be the *omniscient credence function*, *i.e.*, the credence function that centres all of its probability density on the hypothesis that describes the *true* value of $\mathcal{V}$ at w, $\mathcal{V}(w)$ (*e.g.*, the hypothesis that describes the true bias of the coin at w). In addition, it measures the proximity of one credence function p to another q by the *Cramér-von Mises distance* between the two: $\mathfrak{C}(p,q)=\int_{-\infty}^{\infty}|P(x)-Q(x)|^2dx$. The proximity of p to q, on this view, is a function of the area between their *cumulative distri-*

bution functions (CDFs), P and Q (counting regions of smaller divergence between P and Q for less, and regions of greater divergence for more; see figure 2). The CDF P of a distribution p over hypotheses about the value of a variable $\mathcal{V}$ is defined by $P(x) = p(\mathcal{V} \leq x)$, and specifies the probability (according to p) that $\mathcal{V}$ takes a value less than or equal to x.

So to recap, the CRPS says: accuracy at a world w is a matter of *getting close to the truth* at that world, in the sense of pinning down estimates that are close to perfect or vindicated at w, where (i) closeness is measured by Cramér-von Mises distance, and (ii) the (alethically) perfect estimates at w are those given by the omniscient credence function, υ_w, which centres all of its probability density on the hypothesis that describes the true value of $\mathcal{V}$ at w, *viz.*, $\mathcal{V}(w)$.

With these two components in place — a notion of perfection or vindication for credence functions, and a reasonable measure of distance between credence functions — it is easy to see why the CRPS takes the form that it does. We just need one more observation: The cumulative distribution function Υ_w of the perfect, or vindicated distribution υ_w over hypotheses about $\mathcal{V}$ is equal to:

$$\mathbb{1}(x \geq \mathcal{V}(w)) = \begin{cases} 0 \text{ if } x < \mathcal{V}(w) \\ \\ 1 \text{ if } x \geq \mathcal{V}(w) \end{cases}$$

The reason is simple: $\Upsilon_w(x) = \upsilon_w(\mathcal{V} \leq x)$ specifies the probability, according to υ_w, that $\mathcal{V}$ takes a value less than or equal to x. And υ_w assigns probability 1 to the hypothesis that describes the true value of $\mathcal{V}$ at w, *i.e.*, $\mathcal{V}(w)$, and probability 0 to all other hypotheses. This means *inter alia* that it assigns probability 1 to the proposition that $\mathcal{V}$ takes a value less than or equal to x for any x greater than or equal to $\mathcal{V}(w)$. Shorter: $\Upsilon_w(x) = 1$ if $x \geq \mathcal{V}(w)$. Likewise, υ_w assigns probability 0 to the proposition that $\mathcal{V}$ takes a value less than or equal to x for any x strictly less than $\mathcal{V}(w)$. Shorter: $\Upsilon_w(x) = 0$ if $x < \mathcal{V}(w)$.

This gives us the CRPS. The inaccuracy of a continuous credal distribution p at a world w, according to the CRPS, is given by its proximity — measured by Cramér-von Mises distance — to the perfect, or vindicated credal distribution at w, υ_w. And the Cramér-von Mises distance from p to υ_w is just $\mathcal{I}(p,w) = \int_{-\infty}^{\infty} |P(x) - \mathbb{1}(x \geq \mathcal{V}(w))|^2 dx$.

The continuous ranked probability score is an attractive measure of accuracy for a variety of reasons. It is *extensional*: The accuracy of a credal distribution over hypotheses about the value of a variable $\mathcal{V}$ is a function exclusively of (i) the various estimates it encodes — of the truth-values of those hypotheses, and of $\mathcal{V}$ itself — and (ii) the actual values of those quantities.[57] How justified that credal distribution is, in light of some body of evidence, how reliably produced it is, etc., makes no difference to its accuracy, according to the CRPS, except insofar as they help to determine how close its estimates are to the actual values of those quantities. The CRPS is *truth-directed*: Moving your credences uniformly closer to the truth (while remaining coherent) always improves accuracy.[58] It is *strictly*

[57]It is worth noting that two distributions p and q might do equally well estimating the *truth-values* of hypotheses about the value of $\mathcal{V}$, intuitively, and yet have different inaccuracy scores according to the CRPS, in virtue of encoding different estimates for $\mathcal{V}$ itself. For example, let $\mathcal{V}$ be the bias of the coin, and p and q be distributions that spread their probability mass uniformly over the ranges $[0.3, 0.4]$ and $[0.4, 0.5]$, respectively. That is, the density function function that defines p, f_p, attaches the same positive probability mass (=10) to any hypothesis $\mathcal{V} = x$ with $x \in [0.3, 0.4]$, and probability mass 0 to every other hypothesis. Similarly for q (*mutatis mutandis*). Now suppose that the true bias of the coin is 0.6. The value of $\mathcal{V}$ at the actual world w is 0.6. Then p and q seem to do equally well estimating the truth-values of hypotheses about the value of $\mathcal{V}$. They each spread their probability mass uniformly over sets of equal Lebesgue measure — sets containing only false hypotheses — and consequently attach no probability mass to the true hypothesis, $\mathcal{V} = 0.6$. Yet, p's best estimate of $\mathcal{V}$, $Exp_p(\mathcal{V}) = 0.35$, is further from $\mathcal{V}$'s true value, 0.6, than q's best estimate, $Exp_q(\mathcal{V}) = 0.45$. The CRPS treats this fact as relevant to the accuracy of p and q: $\mathcal{I}(p,w) \approx 0.233$ and $\mathcal{I}(q,w) \approx 0.133$.

This is a feature of the CRPS, not a bug. It is precisely to yield accuracy assessments that are "sensitive to distance" in this way — assessments that "give credit" to a distribution for determining a best estimate (expectation) that is close to the true value of the variable by "assigning high probabilities to values near but not identical to the one materializing" — that statisticians sometimes opt for scores like the CRPS, rather than the predictive deviance score, or the ignorance score, etc. (Gneiting and Raftery, 2007, pp. 365-367). This is precisely the reason that statisticians such as Gneiting and Raftery deem it "more compelling to define scoring rules directly in terms of predictive cumulative distribution functions," as the CRPS does, rather than in terms of predictive densities (Gneiting and Raftery, 2007, p. 365).

[58]More carefully, the CRPS is truth-directed in the following sense: if p and q are both probabilistically coherent, continuous distributions over hypotheses about the value of a variable $\mathcal{V}$, and (i) $|p(\mathcal{V} \in S) - w(\mathcal{V} \in S)| \leq |q(\mathcal{V} \in S) - w(\mathcal{V} \in S)|$ for all $S \subseteq \mathbb{R}$, and (ii) $|p(\mathcal{V} \in S') - w(\mathcal{V} \in S')| < |q(\mathcal{V} \in S') - w(\mathcal{V} \in S')|$

proper: Any (coherent) continuous credal distribution expects itself to be more accurate than any other distribution (Gneiting and Raftery, 2007, p. 367).[59] Perhaps most telling, it is a natural extension of the *Brier score*, $\mathfrak{B}(p,w) = (1/n) \cdot \sum_{i=1}^{n} |p(H_i) - w(H_i)|^2$ — a paradigmatically reasonable scoring rule for discrete distributions — to the space of continuous distributions (*cf.* Joyce, 2009, §12, and Gneiting and Raftery, 2007, p. 365).[60] It also yields the correct verdict about comparative accuracy in those cases where obviously correct answers are to be had.[61]

5.3 Case Study Continued

The bookie hands you and your friend the coin. You adopt the uniform prior credal distribution u over hypotheses about the coin's bias B. Your friend adopts a more concentrated prior credal distribution b (figure 3). Now imagine that the true bias of the coin is 5/7. The value of B at the actual world w is 5/7. Imagine also that when you flip the coin your allotted 14 times, it comes up heads 10 times and tails 4. So your new data D is perfectly non-misleading. The true bias of the coin is exactly the frequency of heads in D, *viz.*, 5/7.

When you and your friend condition on your new data D, the two of you arrive at

for some $S' \subseteq \mathbb{R}$, then (iii) $\mathcal{I}(p,w) < \mathcal{I}(q,w)$. This is weaker than the notion of truth-directedness found in (Joyce, 2009, p. 269), which applies unrestrictedly, not simply to coherent distributions. But the CRPS is not defined for incoherent distributions, and so should not be expected to satisfy Joyce's more general truth-directedness constraint.

[59]For more on propriety, which goes under various names in the literature, see the discussion of cogency in (Oddie, 1997, §3), the discussion of self-recommendation in (Greaves and Wallace, 2006, §3), and the discussion of propriety in (Joyce, 2009, §8).

[60]A distribution $p : \mathcal{F} \to [0,1]$ over hypotheses about the values of variables $\mathcal{V}_1, ..., \mathcal{V}_n$ is called discrete when its cumulative distribution function $P : \mathcal{F} \to [0,1]$ is somewhere discontinuous (*cf.* footnote 55 for additional detail). If, for example, you have credences about a finite number of propositions, which can take only two values (1 at worlds where they are true, and 0 where false), then your credal distribution is discrete.

[61]For example, if p and q are beta distributions with the same mean x, but increasing variance, then p is more accurate (less inaccurate) than q at a world w where the true value of $\mathcal{V}$ is x, according to the CRPS: $\mathcal{I}(p,w) < \mathcal{I}(q,w)$. Similarly, if p and q are beta distributions with the same variance, but their means — x and y, respectively — are such that $z < x < y$, then p is more accurate (less inaccurate) than q at a world w' where the true value of $\mathcal{V}$ is z: $\mathcal{I}(p,w') < \mathcal{I}(q,w')$.

the posterior credence functions $u' = u_D$ and $b' = b_D$, respectively (figure 4). The result: your friend is more successful. Her posterior is more accurate than yours: $\mathcal{I}(u', w) = 0.028 > 0.020 = \mathcal{I}(b', w)$. *But her accuracy is less a product of cognitive ability than yours.*

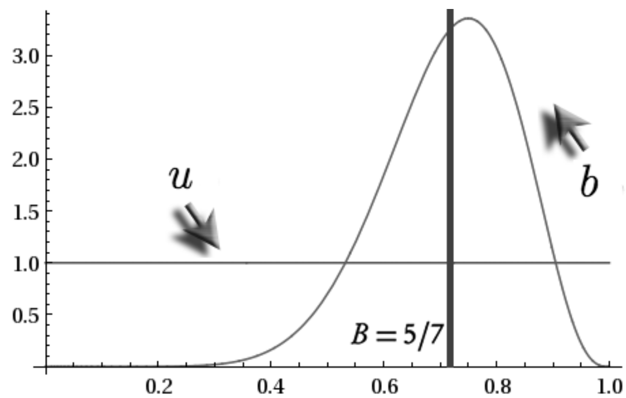

Figure 3: u and b.

The reason, informally, is that your friend's credences are rather resilient with respect to a wide range of data. Not only do they *in fact* remain relatively unchanged by updating on D (evident from figures 3 and 4), but they *would* have remained similarly unchanged, even if the character of her evidence had been a bit different: had she observed 8 or 9 heads, rather than 10, for example.[62] As a result, the accuracy of *her* credences seems to have relatively little to do with the character of her evidence. It is explained primarily by the (fortuitous) accuracy of her *prior* credences. The true bias of the coin, *viz.*, 5/7, happened to be equal to the mean of your friend's rather concentrated prior credence function b. So the hypothesis describing that true bias (5/7) happened to be in the centre of the neigh-

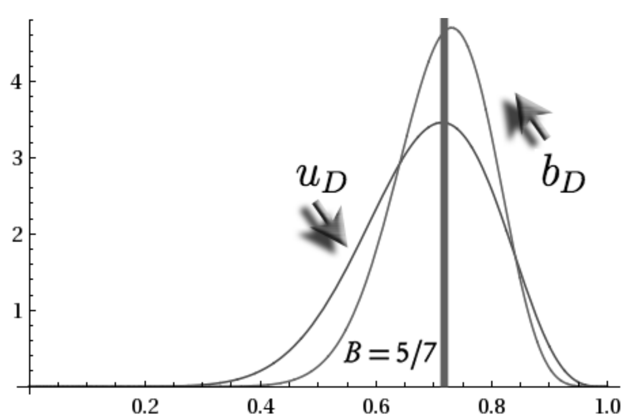

Figure 4: u_D and b_D.

[62]The Cramér-von Mises distance between b and b_D is fairly small: only about 0.002. The Cramér-von Mises distance between u and u_D, in contrast, is 0.068 (34 times bigger). Had your friend observed 8 or 9 heads, rather than 10, the distance between her prior and posterior would have been 0.006 and 0.016, respectively. Yours would have been 0.046 and 0.033 (8 times and 2 times bigger, respectively).

borhood where b lumps nearly all of its probability density. Shorter: her hunch — that the coin's bias is approximately 5/7 — happened to be correct. And her posterior is accurate, in no small part, *because* her prior happened to be so accurate (her hunch was correct). This is constitutive, on our view, of failure to manifest cognitive skill or ability. *Your* degree of posterior accuracy is explained, on the other hand, primarily by the character of your evidence: the fact that it is fairly weighty (14 tosses, rather than only 2 or 3), the fact that it is maximally non-misleading (the frequency of heads in your data sample matches the coin's true bias), and so on.

Our summary statistic — variance in objective expected posterior accuracy across theoretical hypotheses — bears this out. It predicts that the accuracy of your posterior credences is, to a much greater extent than your friend's, the product of cognitive ability. The reason: u's objective expected posterior accuracy varies fairly minimally across hypotheses about the bias of the coin: much less than b's.

To see this, let's simply calculate and compare the objective expected posterior accuracy of u and b, respectively, relative to each hypothesis, $B = x$, about the bias of the coin. Assume that the trials of your experiment — the tosses of the coin — are *independent and identically distributed*. That is, the chance of getting heads on any toss i, given that you get a heads on toss j, is just the unconditional chance of getting a heads on toss i (tosses are independent). And the unconditional chance of getting a heads on any toss i is the same as the unconditional chance of getting a heads on any other toss j (tosses are identically distributed). In that case, each hypothesis, $B = x$, about the bias of the coin fully specifies chances for all of the possible outcomes you might observe: every possible finite sequence of heads and tails. In particular, if the bias B of the coin is x, then the chance of your 14 flips resulting in any sequence with k heads and $14-k$ tails is $x^k \cdot (1-x)^{14-k}$. For example, let S be the proposition $H_1 \& \ldots \& H_k \& T_{k+1} \& \ldots \& T_{14}$ (heads on the first k tosses, tails on the next $14-k$ tosses), and let ch_x be the function that would specify the chances if $B = x$ were true. Then supposing the bias B of the coin is x, the chance of S is:

$$
\begin{aligned}
& ch_x(S) && \\
=& ch_x(H_1 \& ... \& T_{14}) && \\
=& ch_x(H_1 | H_2 \& ... \& T_{14}) \cdot ch_x(H_2 | H_3 \& ... \& T_{14}) \cdot ... \cdot ch_x(T_{14}) && (\textit{Axioms of Probability}) \\
=& ch_x(H_1) \cdot ... \cdot ch_x(T_{14}) && (\textit{Independence}) \\
=& x^k \cdot (1-x)^{14-k} && (\textit{Identically Distributed})
\end{aligned}
$$

And the chance of your 14 flips resulting in *some* sequence or other with k heads and $14-k$ tails — call this proposition D_k — is:

$$ch_x(D_k) = \binom{14}{k} \cdot x^k \cdot (1-x)^{14-k}$$

The propositions $D_0, ..., D_{14}$ are the possible new data items you might receive (and are mutually exclusive and jointly exhaustive). They capture all of the information you might learn — how many times the coin came up heads and tails, respectively — that is relevant to adjudicating between hypotheses about the bias of the coin. (The order of heads and tails, and so on, is irrelevant. The number of heads is what statisticians call a 'sufficient statistic'.)

Now recall that updating (conditioning) your prior credence function, u, on these bits of new evidence yields different posterior credence functions, $u_{D_0}, \ldots, u_{D_{14}}$, each of which is (or at least could be) accurate to a different degree. And since chance has views about which of these data items you will receive, it also has views about how likely you are to attain these different degrees of posterior accuracy.

These views are summed up by u's objective expected posterior accuracy — chance's best estimate of how successful u will be — which we can now calculate as follows:[63]

[63] Line 4 assumes that the coin's bias is not itself a chancy matter, in the following sense: ch_x attaches zero probability to any world w such that $B(w) \neq x$.

$$
\begin{aligned}
& Exp_{ch_x}(\mathcal{I}_{u'}) && (1)\\
&= \sum_{k=0}^{14} \sum_{w \in D_k} ch_x(w) \mathcal{I}(u_{D_k}, w) && (2)\\
&= \sum_{k=0}^{14} \sum_{w \in D_k} ch_x(w) \int_{-\infty}^{\infty} |U_{D_k}(y) - \mathbb{1}(y \geq B(w))|^2 dy && (3)\\
&= \sum_{k=0}^{14} \sum_{w \in D_k} ch_x(w) \int_{-\infty}^{\infty} |U_{D_k}(y) - \mathbb{1}(y \geq x)|^2 dy && (4)\\
&= \sum_{k=0}^{14} ch_x(D_k) \int_{-\infty}^{\infty} |U_{D_k}(y) - \mathbb{1}(y \geq x)|^2 dy && (5)\\
&= \sum_{k=0}^{14} \binom{14}{k} x^k (1-x)^{14-k} \int_{-\infty}^{\infty} |U_{D_k}(y) - \mathbb{1}(y \geq x)|^2 dy && (6)
\end{aligned}
$$

The mathematical details are not terribly important. What is important is that we have a well-motivated, concrete way of measuring chance's best estimate of how successful u will be if the bias B of the coin is x. So we can examine how much this best estimate — u's objective expected posterior accuracy — varies across hypotheses about the coin's bias. And, as it turns out, *it varies fairly minimally*, as figure 5 shows. This tells us something important about what explains the accuracy of your posterior credences, and in turn, the amount of *cognitive ability* that your credal state manifests.

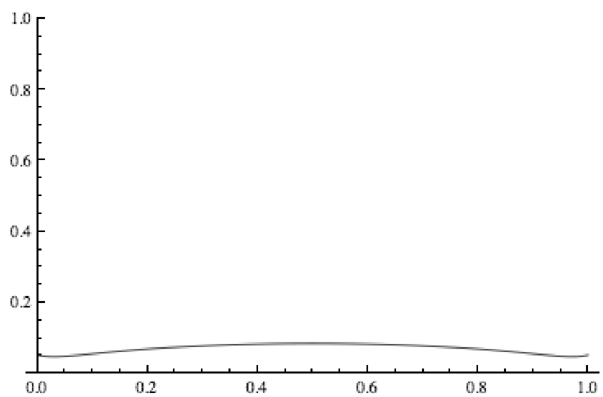

Figure 5: The objective expected posterior accuracy of u across hypotheses $B = x$.

The story — of what explains your success and why — is familiar by now. Suppose chance's best estimate of how successful your credences will be — your objective expected posterior accuracy — would be the same, regardless of how accurate a hunch those credences happen to encode. That is, suppose your prior credence function, u, satisfies:

Invariance. $Exp_{ch_x}(\mathcal{I}_{u'}) = Exp_{ch_y}(\mathcal{I}_{u'})$ for all $0 \leq x, y \leq 1$.

(The uniform prior credence function u does not *perfectly* satisfy INVARIANCE, in the case at hand. Other prior credence functions, as we will see in §6, vary less across hypotheses about the coin's bias. But u does come pretty close: $Exp_{ch_x}(\mathcal{I}_{u'})$ is roughly 0.065 for all $0 \leq x \leq 1$.) Then the accuracy of that hunch is (more or less) irrelevant for explaining why your *chance* of being accurate to any particular degree x, y, or z is what it is. And chances are explanatory foci. So having a spot-on (or wildly-off) hunch also plays no role in explaining why you *actually* attain the degree of posterior accuracy that you do (*viz.*, $\mathcal{I}(u', w) \approx 0.028$). The accuracy of your posterior credences is explained, rather, by the character of your evidence: the fact that it is fairly weighty (14 tosses, rather than only 2 or 3), the fact that it is maximally non-misleading (the frequency of heads in your data, *viz.*, 5/7, matches the coin's true bias), and so on. And this is just what is required for the accuracy of your posterior credences to be the product of cognitive ability.

Your situation mirrors the Mars rover's in many ways. Suppose that chance's best estimate of how successful the rover will be — its objective expected distance from the landing site — would be the same, regardless of whether it happens to emerge from its initial (blind) descent directly above the site, or 1/2 mile to the north, or 3/4 miles to the northeast, etc. (Even the fanciest Mars rover does not *perfectly* satisfy this supposition. But in virtue of the actual Mars rover's skill at landing, it comes pretty close.) Then its initial proximity to the landing site (a non-evidential factor) is (more or less) irrelevant for explaining why its *chance* of landing within 1, or 2, or 3 miles of that site, etc., is what it is. And chances are explanatory foci. So initial proximity is also (more or less) irrelevant for explaining why it attains the *actual* degree of success that it does — why it actually lands close to its target (or not).

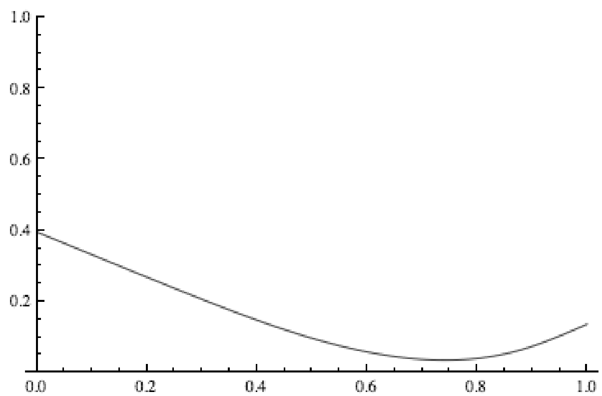

Figure 6: The objective expected posterior accuracy of b across hypotheses $B = x$.

Your friend, however — the one who adopts the more biased beta prior b — is in a different boat. Her prior's objective expected posterior accuracy varies rather significantly across hypotheses about the coin's bias (figure 6). So the fact that she happened to have such a high chance of success (such a low chance of posterior inaccuracy) is explained, in no small part, by the fact that the true bias of the coin, *viz.*, 5/7, happened to be in the centre of the neighborhood where b lumps the majority of its probability density. It is explained, in no small part, by the fact that her hunch — that the coin's bias is approximately 5/7 — happens to be correct. Having a spot-on hunch, then, also plays a significant role in explaining why she *actually* attains the degree of posterior accuracy that she does (*viz.*, $\mathcal{I}(b', w) \approx 0.020$).

Her situation is like a less skilled rover's. A less skilled rover might emerge from its initial descent directly above its landing site. And its landing strategy might give it a high chance of success in that case. So it might land close to its target. Still, unlike the skilled Mars rover, *its* objective expected degree of success varies significantly across changes in initial proximity. How close it happens to be to the landing site after emerging from its initial (blind) descent makes a big difference to its chance of success. The upshot: non-evidential factors (facts about initial proximity) *are* relevant for explaining why its chance of success is what it is. In turn, they *are* relevant for explaining why it is successful to the degree that it is (why it lands close to its target or not).

This all serves to highlight an important virtue of MaxEnt. It delivers posterior credences whose accuracy is, to a large extent, *the product of cognitive ability*. In this respect, it outperforms other brands of objective Bayesianism that recommend more concentrated priors. Unlike those other brands of objective Bayesianism, the accuracy of MaxEnt's posterior credences is *explained* primarily by facts about the evidence that they were conditioned on — how weighty it is, how specific it is, how misleading it is, etc. — rather than, *e.g.*, the fact that its recommended prior credences encode a spot-on (or wildly-off) hunch (a non-evidential factor). And this is just what is required for posterior accuracy to be the product of cognitive ability.

So MaxEnt *does* seem to recommend credal states that are at least *reasonably* eligible candidates for constituting probabilistic knowledge (at least in simple inference problems); the sorts of states in which we might be plausibly said to know, *e.g.*:

(20) that the coin might come up heads on the next toss.

(21) that it's more likely than not that the coin's bias is between 0.6 and 0.8.

(22) that the coin's bias is probably greater than 0.5.

But MaxEnt does not deliver credences that *maximally* eligible candidates for constituting probabilistic knowledge, I will argue. An alternative objectivist method, *the MaxSen method*, delivers better candidates.

6 The MaxSen Method

Once more, suppose that our agent has opinions about propositions in an atomic Boolean algebra $\mathcal{F}$. She designs an experiment to adjudicate between competing theoretical hypotheses, $H_1, \ldots, H_n$ (pairwise incompatible and jointly exhaustive). And her experiment might yield any of the possible data items, $D_1, \ldots, D_m$ (also pairwise incompatible and jointly exhaustive). Suppose that $\mathcal{F}$ contains $H_1, \ldots, H_n$ and $D_1, \ldots, D_m$. Finally, let ch_i be the function that would specify the chances if H_i were true. So the chance that the experiment would yield outcomes $D_1, \ldots, D_m$ if H_i were true is given by $ch_i(D_1), \ldots, ch_i(D_m)$, respectively.

According to the *maximum sensitivity principle*, or *MaxSen*, our agent should take her prior information into account as follows:

- Firstly, she should summarise that prior information by a system of constraints on prior credence functions, $b : \mathcal{F} \to [0,1]$, *e.g.*:

  - $b(H_i) \geq b(H_j)$ (*i.e.*, H_i is at least as probable as H_j)
  - $0.6 \leq b(H_i) \leq 0.9$ (*i.e.*, H_i is probable to at least degree 0.6 and at most 0.9)
  - $b(H_i|D) \geq b(H_j|D')$ (*i.e.*, H_i conditional on D is at least as probable as H_j conditional on D')

  which we represent by a set $\mathcal{C}$ of credence functions, *viz.*, the set of all minimally rational credence functions, b (*i.e.*, b that satisfy Probabilism, the Principal Principle, etc.), consistent with those constraints.

- Secondly, she should adopt some prior credence function $c : \mathcal{F} \to [0,1]$ or other that minimizes

  $$var(c) = max_i \; Exp_{ch_i}(\mathcal{I}_{c'}) \text{ - } min_j \; Exp_{ch_j}(\mathcal{I}_{c'})$$

  over $\mathcal{C}$. (Under a broad range of conditions, this 'MaxSen prior' will turn out to be unique.[64])

[64]Suppose that $\mathcal{C}$ is closed, bounded and convex (*cf.* footnote 48). And suppose that for any f, g and $c = \lambda \cdot f + (1-\lambda) \cdot g$ in $\mathcal{C}$ (where $0 < \lambda < 1$), $\mathcal{I}$ satisfies two constraints, which we might call *worst-case expected convexity* and *best-case expected concavity*:

MaxSen first says that you ought to choose your prior credence function from amongst the minimally rational credence functions that satisfy the constraints imposed by your prior evidence. This bit is common to all kinds of objective Bayesianism, including MaxEnt. Different Bayesians disagree about the nature of one's prior evidential constraints, and provide different justifications for complying with them. They also disagree about which minimally rational credence functions consistent with those prior evidential constraints are rationally permissible *full stop*. But everyone agrees on this much: all minimally rational credence functions obey some fairly uncontroversial epistemic norms. They obey Probabilism, for example, and some version of Lewis' Principal Principle.

These norms, in turn, guarantee that the candidate prior credence functions that you are choosing between — the minimally rational credence function consistent with the constraints imposed by your evidence — are not *too* badly off in terms of accuracy. For example, Joyce (1998, 2009) shows that all and only the probabilistically coherent credence functions avoid the epistemic sin of *accuracy-domination*. Every probabilistically incoherent credence function b is accuracy-dominated by a coherent credence function c, in the sense that b is strictly less accurate than c regardless of which world is actual. No coherent credence function, in contrast, is accuracy-dominated in this way.[65] Similarly,

(*WCE-Convexity*) $max_i\ Exp_{ch_i}(\mathcal{I}_{c'}) < \lambda \cdot [max_j\ Exp_{ch_j}(\mathcal{I}(f'))] + (1-\lambda) \cdot [max_k\ Exp_{ch_k}(\mathcal{I}(g'))]$

(*BCE-Concavity*) $min_i\ Exp_{ch_i}(\mathcal{I}_{c'}) > \lambda \cdot [min_j\ Exp_{ch_j}(\mathcal{I}(f'))] + (1-\lambda) \cdot [min_k\ Exp_{ch_k}(\mathcal{I}(g'))]$

If $\mathcal{I}$ satisfies *BCE-Concaivity*, then whenever c strikes a compromise between f and g (*i.e.*, $c = \lambda \cdot f + (1-\lambda) \cdot g$), the worst case expected inaccuracy of c is better, on balance, than the worst case expected inaccuracies of f and g, respectively. So compromising has an alethic benefit: it safeguards against catastrophic worst-case failure. Similarly, if $\mathcal{I}$ satisfies *WCE-Convexity*, then whenever c strikes a compromise between f and g, the best case of the less compromising f and g is, on balance, better than that of the compromise c. So compromising also has an alethic cost: it forgoes the possibility of maximal best-case success.

When $\mathcal{I}$ satisfies these two constraints (and is continuous), *var* is a continuous, strictly convex, real-valued function. And any such function takes a unique minimum on any closed, bounded, convex set $\mathcal{C}$. Specifying what to do when there is no unique MaxSen prior — when *var* does not take a unique minimum on $\mathcal{C}$, so that MaxSen does not yield a unique recommendation about which prior credence function to adopt — is beyond the purview of our project. The aim, recall, is not to provide *the* definitive set of formal tools for choosing prior credences, but rather, to illustrate one *kind* of tool for securing probabilistic knowledge in a range of inference problems.

[65]See also Predd et al. (2009), Schervish et al. (2009), and Pettigrew (2015).

Pettigrew (2013b, 2015) shows that all and only the credence functions that satisfy the Principal Principle avoid the epistemic sin of *prior chance-domination*. If a credence function b violates the Principal Principle, then it is prior chance-dominated by a credence function c that satisfies it in the following sense: every chance function consistent with your prior evidence expects b to attain a strictly lower degree of prior accuracy than c. No credence function that satisfies the Principal Principle, in contrast, is chance-dominated in this way.

To recap, then, MaxSen comes in two bits. The first directs you to choose a prior credence function from amongst the minimally rational credence functions that satisfy the constraints imposed by your prior evidence. This guarantees that your prior credal state is not *too* badly off in terms of accuracy. At the very least, it avoids the epistemic defects of accuracy-domination and chance-domination, respectively. MaxSen then says that, from amongst the minimally rational credence functions that satisfy the constraints imposed by your prior evidence, you ought to select the prior that minimizes variance in objective expected posterior accuracy across the theoretical hypotheses under investigation. This bit is unique to MaxSen. The motivation for minimizing variance is familiar: If chance's best estimate of how successful your credences will be — your objective expected posterior accuracy — would be the same, or invariant, regardless of how a certain non-evidential factor behaves — fortuitous prior accuracy, in particular — then that factor plausibly plays no role in explaining why your *chance* of being accurate to degree x, y, or z is what it is. And chances are explanatory foci. So it also plays no role in explaining why you *actually* attain the degree of posterior accuracy that you do.

Selecting *prior* credences by minimizing variance, then, will yield *posterior* credences whose accuracy is explained, as much as possible, by evidential rather than non-evidential factors; by the fact that your evidence had a particular character (weighty, specific, non-misleading, etc.), rather than the fact that from amongst *all* of the credal states consistent with your prior evidence, you *happened* to settle on a particularly accurate (or inaccurate) one. And this is just what is required for accuracy to be the product of cognitive ability. So selecting prior credences in this way will yield posterior credences whose accuracy is, to the greatest extent possible, the product of cognitive ability. The upshot: MaxSen's posterior credences will be particularly eligible candidates for constituting probabilistic knowledge — more so than those delivered by MaxEnt, or any other brand of objective Bayesianism. And it is no mystery why. MaxSen simply directs us to choose the prior credences that give rise to those eligible posteriors, whatever they are. So it beats its competitors *by construction*.

MaxSen treats variance in objective expected posterior accuracy across theoretical hypotheses, somewhat crudely, as the difference in its maximum and minimum expected inaccuracies. Of course, this is only one measure of variance. For example, if there were some good epistemic reason to privilege a particular base measure on the space of theoretical hypotheses, for the purposes of integration, then we could substitute the standard statistical measure of variance for our somewhat crude measure.[66] (Though it is difficult to see how we might privilege a particular base measure without simply begging the question in favor of some prior credences or other.) But the crude measure will do, for our purposes. It places an upper bound on statistical variance (whatever base measure you choose), and plausibly any other measure of variance worth its salt. So invariance according to our crude measure entails invariance according to any reasonable measure.

More importantly, our aim is to illustrate one *kind* of formal tool for securing probabilistic knowledge, not to provide *the* definitive set of tools for that end. If successful,

[66]Suppose there is good epistemic reason to privilege a particular base measure μ on the space $\mathcal{H}$ of theoretical hypotheses, for the purposes of integration. Then substituting the standard statistical measure of variance in for our measure (difference between maximum and minimum expected inaccuracy) yields a version of MaxSen that recommends minimizing $Var(c) = \int_{\mathcal{H}} f(c,x)\mu(dx)$, where $f(c,x)$ is the squared difference between the mean objective expected posterior inaccuracy of c (relative to μ), on the one hand, and the expected inaccuracy determined by ch_x (the true chance function if H_x were true), on the other: $f(c,x) = [\int_{\mathcal{H}} Exp_{ch_y}(\mathcal{I}_{c'})\mu(dy) - Exp_{ch_x}(\mathcal{I}_{c'})]^2$.

what we will end up with is a *framework* for theorizing about how to secure probabilistic knowledge. That is what we are really after. So we need not fuss too much about whether our measure of variance is just right.

To illustrate the MaxSen method, imagine once more that a bookie hands you a coin, and offers you a bet. You have no prior evidence about the coin's bias. (Or rather, you have *almost* no prior evidence. To make it manageable to compute the MaxSen prior, assume that your prior evidence rules out any credence function that does not take the form of a beta distribution.[67]) The generous bookie, however, allows you to flip the coin n times before deciding whether or not to take the bet.

MaxEnt recommends that you take your prior information (or lack thereof) into account, in your inference problem, by adopting uniform prior credences over hypotheses about the bias of the coin. It recommends, always and everywhere, adopting credences that are as close to uniform as possible, while satisfying the constraints imposed by your prior evidence. And in this case, the constraints imposed by your evidence are so loose that you can adopt perfectly uniform credences.

MaxSen, in contrast, recommends adopting non-uniform prior credences. Which prior it recommends depends on additional details regarding the set-up of your experiment: in particular, how many times you are going to flip the coin. (See the discussion of the Likelihood Principle in §9.1 for more on MaxSen's sensitivity to experimental set-up.) For example, if you are going to flip the coin 8 times ($n = 8$), then the

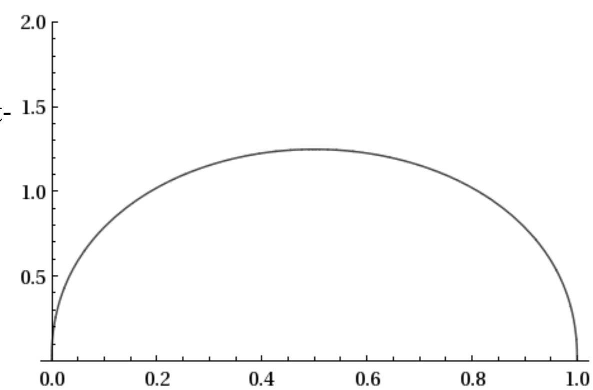

Figure 7: The MaxSen (beta) distribution when $n = 8$ ($\alpha = \beta = 1.45$).

[67]See footnote 53 for more information about beta distributions.

MaxSen prior is the beta distribution with $\alpha = \beta = 1.45$ (figure 7). If instead you are going to flip the coin twenty times ($n = 20$), then the MaxSen prior is the beta distribution with $\alpha = \beta \approx 2$ (figure 8). We will now explore why.

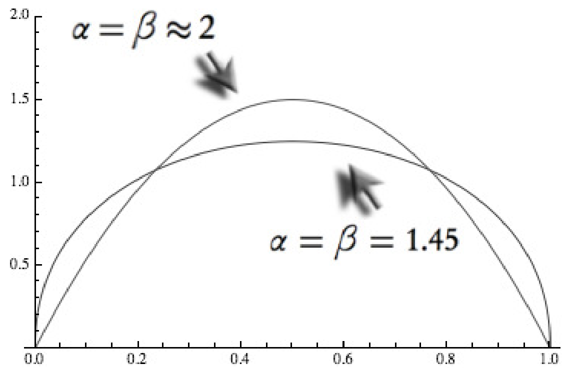

Figure 8: MaxSen beta prior with $n = 8$ ($\alpha = \beta = 1.45$) and $n = 20$ ($\alpha = \beta \approx 2$), respectively.

Suppose that you plan to flip the coin 8 times. To construct the MaxSen prior in this case, we need to find the prior whose objective expected posterior accuracy varies least across hypotheses about the coin's bias. We could use any number of optimization algorithms to do this. But, since our aim is just to illustrate one kind of formal tool for delivering skillfully produced credences, in the hopes of providing a framework for theorizing about how to secure probabilistic knowledge, we will use a rough, but simple technique.

Firstly, we run a simple regression to approximate the variance,

$$var(c) = max_x \; Exp_{ch_x}(\mathcal{I}_{c'}) \text{ - } min_y \; Exp_{ch_y}(\mathcal{I}_{c'}),$$

of every prior credence function c consistent with the constraints imposed by your prior evidence (figure 9). Then we use standard optimization techniques to find the prior for which this approximated variance is smallest. Remember, to make it manageable to compute the MaxSen prior, we assumed that c takes the form of a *beta distribution*. This means that c is fully determined by two *shape parameters*, α and β, which determine how tightly c concentrates its probability density on hypotheses about the coin's bias, of the form $B = x$ (the larger $\alpha + \beta$ is, the more tightly it focuses its density), and where exactly c centres that density. So we can survey the full space of credence functions consistent with your prior evidence by surveying all the values that α and β can take. Inspection of figure 9 reveals that the prior we are searching for — the prior whose objective expected posterior accuracy varies least across hypotheses about the coin's bias — is, approximately, the beta distribution with $\alpha = \beta = 1.4$. Standard optimization techniques show that the prior that minimizes variance (in objective expected posterior accuracy) is, more precisely, the beta distribution with $\alpha = \beta = 1.45$ (figure 7). *This is the MaxSen prior*.

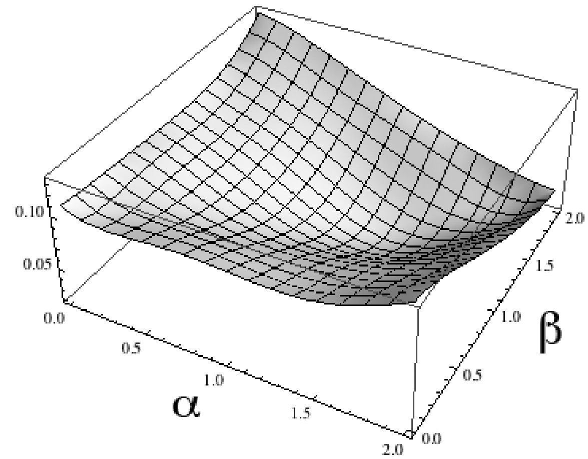

Figure 9: Approximation of *var* with n=8.

In our toy example (when you plan to flip the coin 8 times), MaxEnt delivers credences (uniform credences over hypotheses about the coin's bias) whose posterior accuracy is, to a large extent, the product of cognitive ability. The objective expected posterior accuracy of those credences varies fairly minimally across hypotheses about the coin's bias (figure 5). This is good evidence that their posterior accuracy is explained primarily by facts about the evidence used to update them — how weighty, specific, misleading it is, etc. — rather than fortuitous prior accuracy. And this is what is required for posterior accuracy to be the product of cognitive ability.

All of this notwithstanding, MaxSen delivers posterior credences whose accuracy is, *to the greatest extent possible*, the product of cognitive ability. Their accuracy is explained, to the greatest possible degree, by the character of the evidence on which they are based. The story of why this is so, once again, is familiar.

Suppose chance's best estimate of how successful your MaxSen-recommended prior credences will be — your objective expected posterior accuracy — would be exactly the same, regardless of how accurate a hunch those credences happen to encode. Put differently, suppose it would be the same, regardless of whether the *true* hypothesis about the coin's bias turns out to be $B = x$ — one that your prior credence function s centres a great deal of probability density around (in which case s has a high degree of prior accuracy) — or whether it turns out to be $B = y$ — one that s centres little probability density around (in which case s has a lower degree of prior accuracy). More succinctly, suppose s satisfies:

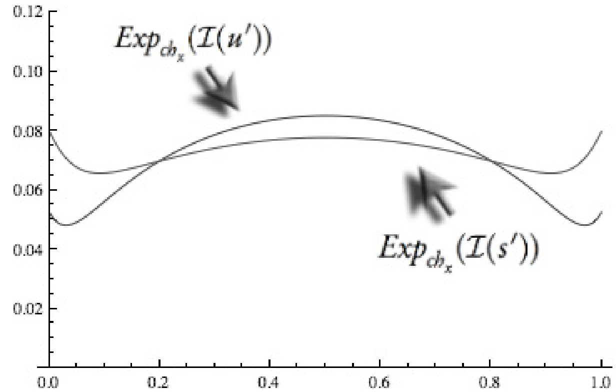

Figure 10: Expected posterior accuracy of both u and s: $Exp_{ch_x}(\mathcal{I}_{u'})$ and $Exp_{ch_x}(\mathcal{I}_{s'})$ (rescaled to emphasize difference in variation across hypotheses $B = x$).

Invariance. $Exp_{ch_x}(\mathcal{I}_{s'}) = Exp_{ch_y}(\mathcal{I}_{s'})$ for all $0 \leq x, y \leq 1$.

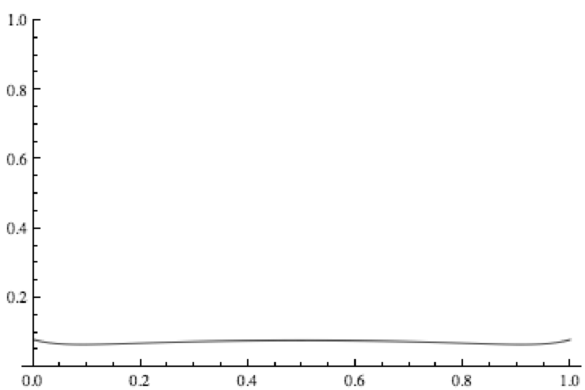

Figure 11: The objective expected posterior accuracy of the MaxSen prior s across hypotheses $B = x$.

(Like the MaxEnt prior u, the MaxSen prior s does not perfectly satisfy this supposition. But it comes as close as any beta distribution can, simply by construction: closer than the credences delivered by MaxEnt (figure 10), or any other brand of objective Bayesianism. And this really is *pretty close* to invariant (figure 11). The difference between s's maximum (=0.077) and minimum (=0.065) objective expected inaccuracies is only 0.012.)

This tells us something important about what explains the accuracy of your posterior credences, and in turn, the amount of cognitive ability that your credal state manifests. Counterfactual independence provides good (if defeasible) evidence for explanatory irrelevance. And the *chance* of your posterior credences attaining any particular degree of accuracy, x, y, or z, is counterfactually independent of how accurate a hunch your prior credences happen to encode. It would be the same, regardless of how accurate that hunch was. (This is what INVARIANCE tells us.) So the accuracy of any such hunch is plausibly irrelevant (or nearly so) for explaining why your *chance* of success (posterior accuracy) is what it is. And chances are explanatory foci. So having a spot-on (or wildly-off) hunch (a non-evidential factor) also plays no role in explaining why you *actually* attain the degree of posterior accuracy that you do. The accuracy of your posterior credences is explained, rather, by the character of your evidence: how weighty, specific, misleading it is, etc. And this is just what is required for posterior accuracy to be the product of cognitive ability.

Once more, the situation is not unlike the skilled Mars rover's. Whether it happens to emerge from its initial (blind) descent directly above the landing site, or 1/2 mile to the north, or 3/4 miles to the northeast, it has roughly the same chance of success (landing close to the target). The result: initial proximity (a non-evidential factor) plays virtually no role in explaining its success. Rather, facts about the quality of its evidence (evidence about its trajectory prior to entering Mars' atmosphere, sensor readings after initial descent, etc.), and how it responds to that evidence are primarily responsible for explaining its success.

You might worry that invariant chances of success come cheap. A defective rover — no parachute, no sensors, no rocket thrusters, etc. — has the same chance of success, regardless of its initial proximity to the landing site, *viz.*, *no* chance! It is bound to be unsuccessful. Similarly, a defective prior — *e.g.*, the incoherent prior c_0 that assigns credence 0 to all competing theoretical hypotheses $H_1, ..., H_n$ and all possible experimental outcomes, or data items $D_1, \ldots, D_m$ — is bound to be unsuccessful (inaccurate). In fact, it is *certain* to attain the same (low) degree of posterior accuracy, on a wide range of plausible inaccuracy measures, come what may.[68] So its *chance* of success is invariant too.

Does MaxSen recommend incoherent priors, then? No! It only appears to do so if we ignore one of MaxSen's two pieces of advice. First, MaxSen says that you ought to choose minimally rational prior credences that satisfy the constraints imposed by your prior evidence. Then it says that your credences should minimize variance (in objective expected posterior accuracy across theoretical hypotheses) from amongst those minimally rational, constraint-satisfying credences. So MaxSen *always* recommends coherent prior credences, so long as minimally rational credence functions are always probabilistically coherent. And all Bayesians agree: they are. Accuracy-dominated credence functions are not even minimally rational. And credence functions avoid the sin of accuracy-domination just in case they are probabilistically coherent.

Still you might worry. Even if a rover is not *fully* defective — perhaps it has a partially functioning parachute, bargain-bin sensors, temperamental rocket thrusters, etc. — it might yet have a *very low* chance of success regardless of its initial proximity to the landing site. As a result, its chance of success might be close to invariant. But such invariance is cheap. There might well be another rover whose chance of success — despite varying more significantly across changes in initial proximity — is *always* higher than that of the partially defective rover.

Similarly, even if a prior is minimally rational, you might worry — even if it obeys Probabilism, the Principal Principle, etc. — it could nevertheless have a uniformly low

[68] More carefully, $c_0 : \mathcal{F} \to [0, 1]$ will attain exactly the same degree of posterior (and prior) accuracy, come what may, relative to any *extensional* inaccuracy measure $\mathcal{I}$. $\mathcal{I}$ is extensional iff $\mathcal{I}(c, w)$ is a function exclusively of $c(X)$ (c's estimate of X's truth value), and $w(X)$ (X's truth-value at w), for all $X \in \mathcal{F}$. Since c_0 assigns exactly the same credence (*viz.*, 0) to every $H_i, D_j \in \mathcal{F}$, and exactly one of element of $\{H_1, \ldots, H_n\}$ and $\{D_1, \ldots, D_m\}$ is true, come what may (both are partitions), c_0's inaccuracy is the same, come what may, according to any extensional $\mathcal{I}$.

objective expected posterior accuracy (low across all theoretical hypotheses). As a result, its objective expected posterior accuracy might be close to invariant. But this sort of invariance seems cheap. For all we have said, there might be another prior whose objective expected posterior accuracy — despite varying more significantly across theoretical hypotheses — is *always* higher. That is, there might be another prior that *posterior chance-dominates* it.

Does MaxSen recommend such defective priors? No! As it turns out, no prior credence function that satisfies the Principal Principle is posterior chance-dominated. We mentioned earlier that Pettigrew (2013b, 2015) establishes the following result: if a credence function satisfies the Principal Principle, then it is not *prior* chance-dominated. That is, no other credence function has uniformly higher objective expected *prior* accuracy (higher across all theoretical hypotheses). In fact, though, something stronger is true: if a credence function satisfies the Principal Principle, then not only is it not *prior* chance-dominated, but it is not *posterior* chance-dominated either. There is no other credence function that has uniformly higher objective expected *posterior* accuracy.[69] So MaxSen never recommends such defective priors, so long as minimally rational credence functions always obey the Principal Principle.

The take-home lesson is this: being *accurate* and being *skillfully produced* are both

[69]To see this, note that if a credence function c satisfies the version of the Principal Principle we introduced earlier (footnote 41) — the extended temporal principle — then $c(X|T_{ch}) = ch(X)$ whenever $c(T_{ch}) > 0$. This means that

$$Exp_c(\mathcal{I}_{b'}) = \sum_{1\leq i\leq m}\sum_{w\in D_j} c(w)\cdot\mathcal{I}(b_{D_j}, w) = \sum_{ch} c(T_{ch})[\sum_{1\leq i\leq m}\sum_{w\in D_j} ch(w)\cdot\mathcal{I}(b_{D_j}, w)]$$

Greaves and Wallace (2006) show that $Exp_c(\mathcal{I}_{b'})$ takes a minimum at $b = c$, so long as $\mathcal{I}$ is a strictly proper inaccuracy score. But now suppose (for reductio) that some credence function $b \neq c$ has uniformly lower objective expected posterior inaccuracy, so that

$$Exp_{ch}(\mathcal{I}_{b'}) = \sum_{1\leq i\leq m}\sum_{w\in D_j} ch(w)\cdot\mathcal{I}(b_{D_j}, w) < \sum_{1\leq i\leq m}\sum_{w\in D_j} ch(w)\cdot\mathcal{I}(c_{D_j}, w) = Exp_{ch}(\mathcal{I}_{c'})$$

for every chance function ch. Then whatever values the $c(T_{ch})$'s take, we have $Exp_c(\mathcal{I}_{b'}) < Exp_c(\mathcal{I}_{c'})$, which contradicts the claim that $Exp_c(\mathcal{I}_{b'})$ takes a minimum at $b = c$.

epistemically laudable properties. Both are plausibly necessary for probabilistic knowledge. And MaxSen is designed to secure both. It directs you to choose minimally rational prior credences that satisfy the constraints imposed by your prior evidence in order to ensure that they are reasonably accurate. At the very least, such credences avoid the epistemic sins of accuracy-domination and chance-domination (prior and posterior). MaxSen then directs you to choose prior credences that minimize variance in objective expected posterior accuracy (from amongst the minimally rational credal states that satisfy those constraints) in order to ensure that they are the product of cognitive ability.

To sum up, if what we have said is correct, then MaxSen delivers credences that are not only reasonably likely to be accurate, in a wide range of inference problems, but whose accuracy is, to the greatest extent possible, explained by facts about the evidence (prior and experimental) on which they are based. In turn, the accuracy of those credences is, as much as possible, the product of cognitive ability.

So MaxSen takes us some way toward securing probabilistic knowledge. It delivers credences that are maximally eligible candidates for constituting such knowledge. It does as much as possible — more than MaxEnt, or any other brand of objective Bayesianism — to help secure a credal state in which you might plausibly be said to know, *e.g.*:

(23) that the next trial of your experiment will almost certainly not produce data D.

(24) that it's more likely than not that the true theoretical hypothesis is either H_i or H_j.

(25) that the true hypothesis is probably one of $H_1, ..., H_k$.

Before concluding, it is worth examining these 'maximally eligible candidates' a bit more closely. They obey a particular *principle of equality*, as we will see in §7. More specifically, MaxSen delivers credences that give theoretical hypotheses equal *consideration*, in a certain sense, rather than equal *treatment*. This tells us something important about the nature of cognitive ability. The demands of cognitive ability manifest themselves as demands of equal consideration.

7 Two Principles of Equality

Peter Singer (1975) argued that the most basic principle of equality is a principle of *equal consideration*. The interests of all — human or otherwise — deserve equal consideration in moral deliberation. Importantly, though, "equal consideration for different beings may lead to different treatment" (Singer 1975, p. 2). For example:

> If I were to confine a herd of cows within the boundaries of the county of, say, Devon, I do not think I would be doing them any harm at all; if, on the other hand, I were to take a group of people and restrict them to the same county, I am sure many would protest that I had harmed them considerably, even if they were allowed to bring their families and friends, and notwithstanding the many undoubted attractions of that particular county. (Singer 1985, p. 6)

Humans have interests "in mountain-climbing and skiing, in seeing the world and in sampling foreign cultures;" whereas cows do not (Singer, 1985, p. 6). So we are not required to treat humans and cows equally, simply in virtue of giving equal consideration to their respective interests. Indeed, we will often be required *not* to treat them equally.

We can and must, then, distinguish between two importantly different principles of equality, according to Singer — principles of equal *treatment* and *consideration*, respectively — which govern our practical and evaluative attitudes toward individuals. Similarly, we can (and, I think, must) distinguish between importantly different principles of equality governing our *doxastic* attitudes toward *theoretical hypotheses*. On the one hand, we have a principle of equal doxastic treatment, which demands that we treat theoretical hypotheses equally by giving them equal credence. On the other hand, we have a principle of equal doxastic consideration, which demands that we reason about theoretical hypotheses in such a way as to give each one an equal chance of being discovered, *i.e.*, of you learning of its truth (or having accurate posterior credences about its truth), if indeed it is true.

If we had explored MaxEnt, and then called it a day before discovering MaxSen, we might have suggested the following: While morality requires us to give individuals equal consideration, the pursuit of probabilistic knowledge requires us to give theoretical hypotheses equal treatment. After all, MaxEnt recommends that you have credences that are as close to uniform, or equal as possible, while satisfying the constraints imposed by your prior evidence. And it seems to deliver credences that are, to a large extent, the product of cognitive ability, and hence good candidates for probabilistic knowledge. So we seem to have learned something surprising about the nature of cognitive ability, and in turn, about the nature of probabilistic knowledge: both involve treating hypotheses equally, by giving them equal credence (or as close to equal credence as possible, while satisfying the constraints imposed by your prior evidence).

But this would have been a mistake. Just as morality requires us to give individuals equal *consideration*, so too does the pursuit of probabilistic knowledge require us to give theoretical hypotheses equal *consideration*. Treating cows and humans equally is *preferable*, from the moral perspective, to the status quo (factory farms and the like). But it does not follow that it is the morally *best* option. It does not follow that morality requires equal treatment. Similarly, treating all theoretical hypotheses equally, by adopting uniform or equal prior credences, is *preferable*, from the epistemic perspective, to adopting various more concentrated priors. It is a better means to the end of securing skillfully produced credences, and in turn, credences that are good candidates for probabilistic knowledge. But it does not follow that it is the *best* option. It does not follow that having such ability or knowledge *requires* equal treatment.

MaxSen, not MaxEnt, delivers the most skillfully produced credences possible, and in turn, maximally eligible candidates for probabilistic knowledge. But MaxSen does not recommend giving theoretical hypotheses equal treatment by giving each one equal credence. Instead, MaxSen recommends giving them equal *consideration* by giving each one the same chance of being discovered, *i.e.*, of you learning of its truth (or having accurate posterior credences about its truth), if indeed it is true. This is just what minimizing variance in objective expected posterior accuracy amounts to.

The primary difference between these two doxastic equality principles — and, in turn, between MaxEnt and MaxSen — is that the principle of equal consideration recognizes that certain theoretical hypotheses speak up more clearly and forcefully than others do, to put it metaphorically. Certain hypotheses have a high chance of producing rather extreme, probative data: data that forces most priors (save for fantastically pigheaded ones) to centre most of their probability density around *them*.[70] They scream, if you will, "I'm true!" This sort of boisterous hypothesis does not need equal treatment (credence) to receive equal consideration. It does not need equal treatment to have the same chance of being discovered as other hypotheses, *i.e.*, of you learning of its truth, if it is true.

For example, if you are going to flip the coin of unknown bias 8 times, MaxSen recommends giving less credence to hypotheses according to which the bias of the coin is closer to 0 or 1, and more credence to hypotheses according to which it is closer to 1/2, as figure 12 reminds us. The reason: hypotheses according to which the bias is close to 0 or 1 speak up more clearly than hypotheses according to which it is close to 1/2. They have a high chance of producing extreme, probative data (almost all tails, or all heads, respectively):

[70]Strictly speaking, of course, hypotheses do not produce data. We ought to say: if certain 'boisterous' hypotheses were true (*e.g.*, if the bias of the coin were 0.99), then there would be a high chance that conducting the experiment (*e.g.*, flipping the coin n times) would produce rather extreme, probative data (*e.g.*, a string of nearly n heads).

data that forces most priors (the MaxSen prior included) to centre most of their probability density around them. So they do not need as much credence (equal treatment) to receive equal consideration.

This is a bit like a teacher in a classroom filled with children, some of whom are boisterous and some of whom are reserved. To give equal consideration to the opinions of all children, the teacher need not pay equal attention to them all. Indeed, she should strive to listen more carefully to the reserved, quiet children: eliciting their thoughts, boosting their confidence, and so on. The boisterous children will make themselves known without such special attention.

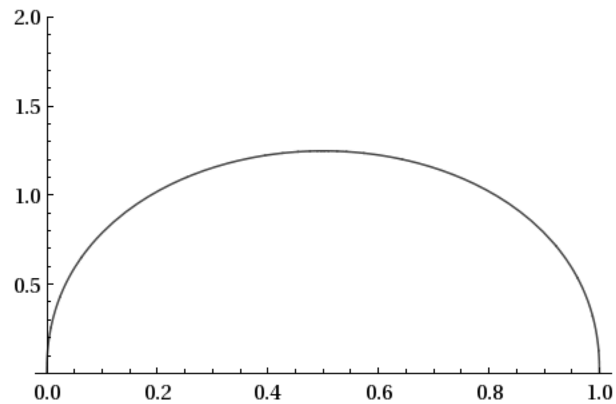

Figure 12: MaxSen prior ($n = 8$).

If this is right, then we *have* learned something surprising about the nature of cognitive ability or skill, and in turn, about the nature of probabilistic knowledge. Having cognitive ability does not involve treating hypotheses equally. Rather, it involves giving them equal consideration. It involves reasoning about them in such a way that you have an equal chance of discovering them (of learning that they are true, if they are).

8 Conclusion

Moss (2013) provides compelling reasons to countenance probabilistic knowledge. Credences can amount to knowledge in much the way that full beliefs can. I suggested that whatever else it takes for an agent's credences to amount to knowledge, they must satisfy both an anti-luck condition and an ability condition: their accuracy must not be lucky (it must be *safe*); and it must be the product of *cognitive ability* or *skill*.

So I set out to find formal tools for choosing prior credences, at least in a range of inference problems, to help ensure those conditions are met. The best tools for this end, I argued, would be designed specifically to yield credences that manifest *cognitive skill*. Normally, cognitive skill mitigates dependence on luck. When it does not, savvy prior construction won't help us much anyway.

I then provided an account of cognitive skill or ability that says: having cognitive ability is a matter of reasoning in such a way that your evidence explains the accuracy (or inaccuracy) of your credences to the greatest degree possible. To help sift credal states that satisfy the ability condition on probabilistic knowledge from ones that do not, I searched for a summary statistic that tracks the amount of cognitive ability that a credal state manifests. I pointed to one statistic that seems to fit the bill: variance in objective expected posterior accuracy across theoretical hypotheses. The smaller this statistic, I argued, the greater the extent to which a credal state's posterior accuracy is a product of cognitive ability.

I then used this statistic to develop a novel kind of objective Bayesianism, MaxSen. MaxSen advises adopting the prior credal state (from amongst the credal states consistent with your prior evidence) that minimizes this statistic. So the posterior accuracy of its credences are, to the greatest extent possible, the product of cognitive ability. Those credences, then, are particularly eligible candidates for constituting probabilistic knowledge.

Finally, I argued that MaxSen teaches us something important about the nature of cognitive ability and probabilistic knowledge. Having credences that amount to probabilistic knowledge, and so *inter alia* are the product of cognitive ability, requires giving theoretical hypotheses equal consideration, rather than equal treatment.

In appendices A and B, I tie up some important loose ends. Appendix A addresses two pressing objections. For example, Venn (1866), Keynes (1921) and Fisher (1922) provide examples that seem to show that MaxEnt yields inconsistent results in a range of cases, depending on how you describe them. The proponent of MaxSen must show that these problems do not extend to her proposal. Appendix B discusses MaxSen's limitations.

9 Appendix A: Objections

9.1 Likelihood Principle

The MaxSen method seems to run afoul of the Likelihood Principle (Edwards et al., 1963, p. 237):

Likelihood Principle. For any two experiments aimed at adjudicating between theoretical hypotheses $H_1, ..., H_n$, and any two items of data D and D' produced by those experiments, if D and D' have the same inverse probabilities (or 'likelihoods'), up to an arbitrary positive constant $k > 0$ — which is just to say that all hypotheses H_i and H_j (with $1 \leq i, j \leq n$) render the first datum, D, k times as probable as the second datum, D': $p(D|H_i)/p(D'|H_i) = k = p(D|H_j)/p(D'|H_j)$ — then the evidential import of D and D' for $H_1, ..., H_n$ is the same.

Many Bayesian statisticians, such as Savage, de Finetti and Berger (as well as frequentist statisticians such as Fisher) take the LP to be central to rational inductive inference. Birnbaum (1962) summarizes the standard Bayesian rationale for the LP as follows. First, on the Bayesian view, according to Birnbaum, the aim of rational inductive inference is to use "experimental results along with other available [prior] information" to determine a posterior that provides "an appropriate final synthesis of available information" (Birnbaum, 1962, p. 299).

Second, Bayes' theorem tells us that posterior (post-experiment) probabilities, $p(\cdot|D)$, are fully determined by two components: a prior (pre-experiment) probability function,

$p(\cdot)$, which specifies how probable the various theoretical hypotheses $H_1, \ldots, H_n$ are in light of and an inverse probability function, or likelihood function, $p(D|\cdot)$, which specifies how probable the hypotheses $H_1, \ldots, H_n$ render D.

Bayes' Theorem. $p(H_i|D) = [p(D|H_i) \cdot p(H_i)]/p(D)$

$$= [p(D|H_i) \cdot p(H_i)]/\textstyle\sum_j p(D|H_j) \cdot p(H_j)$$

Finally, because the *prior* distribution captures the evidential import of the *prior* data (no more, no less), the likelihood function (inverse probabilities) must capture the evidential import of the experimental data, on the Bayesian view (no more, no less). “In this sense,” Birnbaum says, “we may say that [Bayes' theorem] implies [the likelihood principle]” (Birnbaum, 1962, p. 299). “The contribution of experimental results to the determination of posterior probabilities is always characterized just by the likelihood function and is otherwise independent of the structure of an experiment” (*ibid*.).

MaxSen seems to violate the LP by making ‘extraneous’ features of the experimental set-up — in particular, its *stopping rule* — relevant to the evidential import of the experimental data. Stopping rules are rules that specify when to stop gathering new data. For example, a *fixed stopping rule* might tell you to flip a coin of unknown bias exactly 5, or 10, or 15 times. An *optional stopping rule*, on the other hand, might tell you to flip the coin until at least half of the tosses come up heads. Stopping rules have no bearing on the import of the experimental data, according to the LP, because they have no influence on likelihoods. Imagine, for example, that you perform some experiment and it produces some outcome. Whether you plan to perform 5, or 10, or an indefinite number of trials after this first trial, in accordance with various different stopping rules, has no influence on likelihoods. It makes no difference to how probable any of the various theoretical hypotheses under consideration render the outcome of the first trial. So it should have no bearing on the evidential import of that outcome, according to the LP. All such features of the experimental set-up are extraneous.

The argument that MaxSen violates the LP, by making stopping rules relevant to evidential import or force, goes as follows. First, as any proponent of the method would happily admit, stopping rules *are* relevant to which prior you ought to adopt, according to MaxSen. Suppose that you and your friend are going to flip a coin, in order to adjudicate between competing hypotheses about its bias. You adopt different fixed stopping rules: you plan to flip the coin 8 times, while your friend plans to flip it 20 times. Then MaxSen recommends that you adopt a certain beta prior ($\alpha = \beta = 1.45$) which gives more credence to hypotheses according to which the bias is close to $1/2$, and less credence to those according to which it is closer to 0 or 1. It recommends that your friend adopt a different, more concentrated prior ($\alpha = \beta \approx 2$). (See figure 8.)

The argument continues:

1. Posterior probabilities reflect the evidential import of the total data (prior and experimental) for the theoretical hypotheses under investigation (no more, no less).

2. MaxSen renders posteriors sensitive to stopping rules (by rendering priors sensitive to stopping rules).

3. So, according to MaxSen, the evidential import of the total available data is sensitive to stopping rules. (*From 1* & *2*)

4. Stopping rules are obviously irrelevant to the evidential import of the *prior* data. If they are relevant to the import of the total data at all, it must be because they impact the import of the *experimental* data.

5. Hence, the evidential import of the experimental data is sensitive to stopping rules, according to MaxSen. (*From 3* & *4*)

6. The LP says: stopping rules are irrelevant to the evidential import of the experimental data.

C. MaxSen violates the LP. (*From 5 & 6*)

In fact, though, MaxSen is perfectly consistent with the LP. The problem with this argument: premise 1 is false. A rational agent's prior credences reflect more than just the evidential import of her prior data, and her posterior credences reflect more than just the evidential import of her total data. A rational agent engages in inquiry to secure *probabilistic knowledge*, according to the proponent of MaxSen. Her prior credences, then, must also reflect the right sort of *inductive strategy* — the right sort of strategy for handling new data — so that her *posterior* credences are eligible candidates for constituting such knowledge. This means, *inter alia*, encoding an inductive strategy that grounds *cognitive skill* or *ability*.

MaxSen does not render priors sensitive to stopping rules because the evidential import of the experimental data is sensitive to stopping rules. It is not. Rather, it renders priors sensitive to stopping rules because facts about which inductive strategy is best suited to ground cognitive skill or ability are sensitive to stopping rules.

Compare: you are part of the engineering team designing the new Mars rover. You want to know: Which landing strategy makes the rover's chance of success (landing close to the target) more or less invariant, regardless of whether it emerges from its initial (blind) descent directly above the landing site, or 1/2 mile to the north, or 3/4 miles to the northeast, etc.? Which strategy, as a result, makes the rover's initial proximity to the landing site (within reasonable bounds) more or less irrelevant for explaining its success? Clearly, the answer depends on a host of variables, *e.g.*, how much fuel will be available for maneuvering once the rover is finally descending at a safe speed. If quite a bit will be available, *e.g.*, 1000 lbs, then perhaps a long burn in the direction of the landing site, followed by a series of short burns would be best. If much less fuel will be available, *e.g.*, 600 lbs, a long burn might be out of the question.

Similarly, if you and your friend are going to flip a coin, in order to adjudicate between competing hypotheses about its bias, the right question to ask (if you hope to secure probabilistic knowledge) is this: Which inductive strategy (prior credences) makes your chance of success (posterior accuracy) more or less invariant, regardless of how close the true bias happens to fall your prior estimate? Which strategy, as a result, makes fortuitous prior success (accuracy) more or less irrelevant for explaining posterior success (accuracy)? The answer depends on various features of the experimental set-up — in particular, the stopping rule in play. The reason: the stopping rule determines *how much data will be available to 'maneuver' your credences toward the truth*. And just as the best landing strategies given 600 and 1000 lbs of total available fuel, respectively, might recommend using the first 500 lbs differently, so too might the best inductive strategies in different experimental contexts recommend using some initial bit of data differently, if there are different total amounts of experimental data available in those contexts. If you are going to flip the coin 8 times, and your friend is going to flip the coin 20 times — so that your friend has more (weightier) data available to 'maneuver' her credences toward the truth than you do — your best inductive strategies, respectively, might recommend using some initial bit of data (*e.g.*, the outcomes of the first 5 flips) differently (*i.e.*, might recommend drawing different inferences on the basis of this initial bit of data).

The moral: MaxSen is consistent with the LP, despite its sensitivity to stopping rules. Stopping rules *do* determine certain features of the MaxSen prior. But this is not because the evidential import of the experimental data is sensitive to stopping rules, according to MaxSen. It is not. The reason, instead, is that which inductive strategy is best suited to ground cognitive skill or ability in a given context of inquiry is partially determined by which stopping rule is in play that context.

9.2 Language Dependence

As frequentist statisticians and dyed-in-the-wool subjective Bayesians often note, MaxEnt seems to yield inconsistent results in a range of cases, depending on how you describe them. The following example, adapted from (Fisher, 1922, pp. 324-5), illustrates the point. Imagine once more that you are handed a coin. You plan to flip it n times, in order to adjudicate between competing hypotheses about its bias. You have no relevant prior evidence. (Almost no prior evidence, anyway. Once again, to make it manageable to compute the MaxSen prior, assume that your prior evidence rules out any credence function that does not take the form of a beta distribution.)

Given your dearth of evidence, MaxEnt recommends adopting uniform prior credences u over hypotheses about the bias B of the coin. It recommends spreading your credence evenly over all of those hypotheses. But, Fisher points out, you "might never have happened to direct [your] attention to the particular quantity" B (Fisher, 1922, p. 325). You might have focused, instead, on the quantity $\theta = \sqrt{B}$ (or any number of other quantities). "The quantity, θ," Fisher says, "measures the degree of probability, just as well as $[B]$, and is even, for some purposes, the more suitable variable" (*ibid*). To put the point slightly differently, you might have described your inference problem in different terms. not as a problem that requires sorting out which of the various hypotheses about the coin's *bias* B is true, but rather, as one that requires sorting out which hypothesis about the *square root of its bias* is true (what the true value of $\theta = \sqrt{B}$ is), or the *cosine of its bias* ($\delta = \cos B$), etc.

These redescriptions of your inference problem are equivalent, in a sense. The hypotheses that they employ (about the bias of the coin, about the square root of its bias, etc.) divide up logical space in exactly the same way (they partition the space $\mathcal{W}$ of possible worlds into exactly the same sets).[71] And there is no obvious epistemic reason to prefer one way of describing your inference problem over any other. Yet MaxEnt's recommendation differs, depending on which description you choose. For example, if you had focused on the square root of the coin's bias $\theta = \sqrt{B}$, rather than its bias B — if you had framed your inference problem as one that requires sorting out the true value of θ, rather than B — then MaxEnt would have recommended spreading your credence evenly over hypotheses about the value of θ. It would have recommended adopting uniform prior credences u^* over hypotheses about the square root of the coin's bias, which is equivalent to adopting a *non*-uniform prior b^* over hypotheses about its bias, as figure 13 shows.[72]

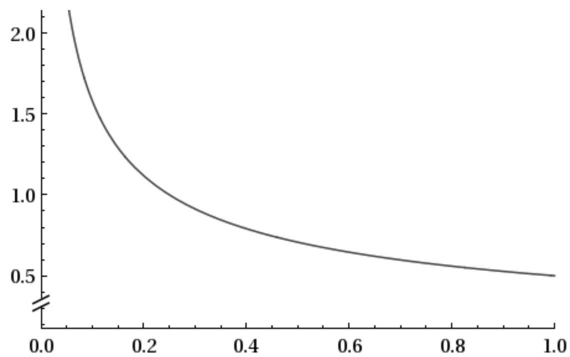

Figure 13: Non-uniform prior b^* over hypotheses about the value of B, which is equivalent to the uniform prior u^* over hypotheses about the value of θ.

So MaxEnt recommends different prior credences, depending on how you describe your inference problem. As a result, it recommends making different (inconsistent) judgments, in those different cases. For example, if you adopt the MaxEnt (uniform) prior u over hypotheses about the coin's bias, B, and then observe two heads in a row, your new best estimate of that bias is 0.75. In contrast, if you adopt the MaxEnt (uniform) prior u^*

[71]For example, for any hypotheses $B = x$ about the bias of the coin, the set of worlds that make the hypothesis $B = x$ true is exactly the set of worlds that makes $\theta = \sqrt{x}$ true (the square root of the coin's bias is $\sqrt{x}$). So the partition $\{\{w \in \mathcal{W} | B{=}x \textit{ is true at } w\} | x \in [0,1]\}$ of $\mathcal{W}$ is equal to $\{\{w \in \mathcal{W} | \theta{=}y \textit{ is true at } w\} | y \in [0,1]\}$.

[72]The uniform prior over hypotheses of the form $\theta = x$ is equivalent to the non-uniform prior over hypotheses of the form $B = x$, defined by the probability density $f(x) = 1/(2\sqrt{x})$ (figure 13).

over hypotheses about the *square root* of the coin's bias, θ, and observe two heads, your best estimate is 0.71.

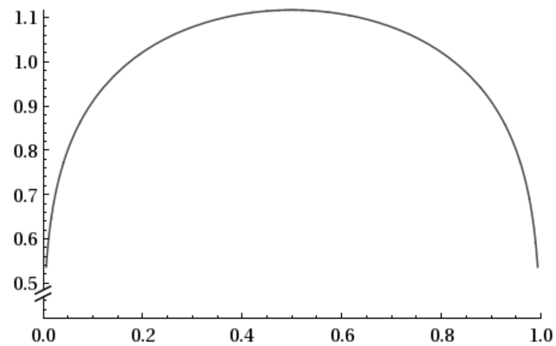

Figure 14: MaxSen prior s for B.

One might suspect that the MaxSen method is subject to a similar sort of language or description dependence. And it is. But MaxSen also furnishes a clear rationale for favoring one description of your inference problem over others. So it is not inconsistent, in the way MaxEnt seems to be. It does not leave you in the precarious position of yielding different prescriptions relative to different ways of describing your inference problem, with no good reason to choose between them.

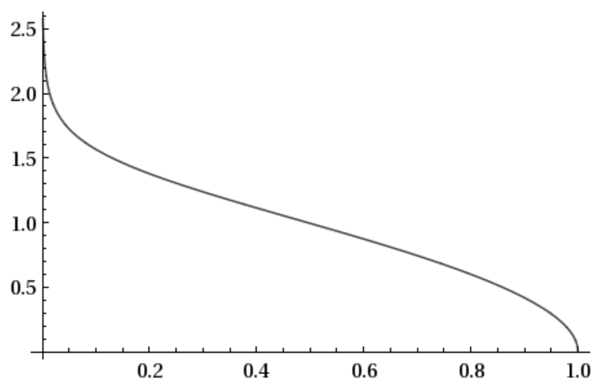

Figure 15: MaxSen prior s^* for θ.

Suppose that you plan to flip the coin of unknown bias 5 times. The MaxSen prior over hypotheses about the coin's bias, B, is the beta distribution s with $\alpha = \beta \approx 1.2$ (figure 14). If, in contrast, you use MaxSen to determine prior credences over hypotheses about the *square root* of the coin's bias, θ, you will arrive at the distribution s^* with $\alpha \approx 0.9$ and $\beta \approx 1.5$ (figure 15). And s^* is *not* equivalent to s. Adopting the prior credence function s^* over hypotheses about the square root of the coin's bias, θ, is equivalent to adopting a rather concentrated, resilient prior c^* over hypotheses about its bias, B; one which favors hypotheses according to which the coin's bias is close to 0 (figure 16).[73]

[73]The MaxSen prior over hypotheses of the form $\theta = x$ is equivalent to the non-symmetric prior over hypotheses of the form $B = x$, defined by the probability density $g(x) = 0.65581\sqrt{1-\sqrt{x}}/x^{0.55}$ (figure 16).

So depending on how you describe your inference problem, MaxSen recommends different prior credences. But this does not amount to inconsistency, because MaxSen furnishes a clear rationale for framing your inference problem as one about B, rather than θ (or any other quantity). It furnishes a clear rationale for describing it as a problem that requires sorting out the coin's true *bias*, rather than the square root of its bias is true, or the cosine of its bias, etc.

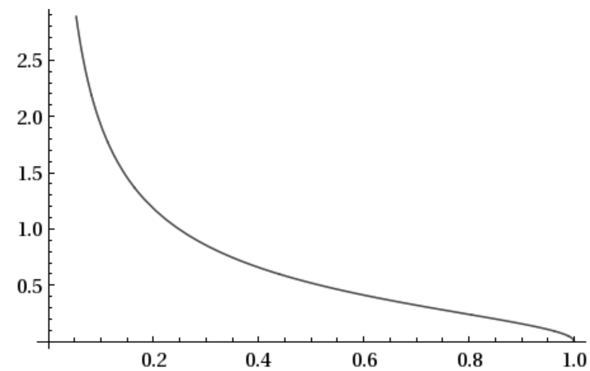

Figure 16: The biased prior c^* over hypotheses about the value of B, which is equivalent to the MaxSen prior s^* over hypotheses about the value of θ.

Rational agents engage in inquiry, the thought goes, with the aim of securing probabilistic knowledge *not simply about theoretical hypotheses* (hypotheses about a virus' infection mechanism, about the causal underpinnings of the climate system, etc.), *but also about non-theoretical propositions* (whether a patient's cancer will remain in remission, whether atmospheric CO_2 levels will double by 2050, etc.).[74] In our example, you aim to not only secure probabilistic knowledge about bias of the coin, or equivalently, the *chance* of heads, but also about whether the coin will *in fact* come up heads on the next toss.

Imagine that you adopt the MaxSen prior s over hypotheses about the bias of the coin, B. Your friend adopts the MaxSen prior s^* over hypotheses about the square root of the coin's bias, θ. You observe some data D and update. Then, if MaxSen lives up to its billing, the accuracy of your posterior credences about B are, to the greatest extent possible, the product of cognitive ability. Those credences, then, are good candidates for

[74]Non-theoretical propositions, for our purposes, are just the propositions to which theoretical hypotheses (chance hypotheses, causal models, etc.) assign probabilities.

probabilistic knowledge.

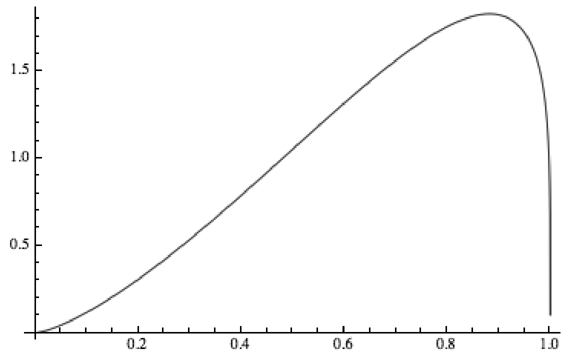

Figure 17: The biased prior c over hypotheses about the value of θ, which is equivalent to the MaxSen prior s over hypotheses about the value of B.

Not so, however, for your credences about θ. Adopting the MaxSen prior s over hypotheses about B is equivalent to adopting a rather concentrated, resilient prior c over hypotheses about θ; one which favors hypotheses according to which θ is close 1 (figure 17).[75] And the objective expected posterior accuracy of c varies fairly *significantly* across hypotheses about the square root of the coin's bias. So the accuracy (or inaccuracy) of your actual posterior credences about θ is explained to a much greater extent by their fortuitous prior accuracy (or inaccuracy), and hence is to a much lesser extent the product of cognitive ability. Similar remarks apply to your friend's credences about B and θ (*mutatis mutandis*).

Both of you seem on par, then. After all, probabilistic knowledge about the chance of heads (which you plausibly have, and your friend lacks) is no more intrinsically epistemically valuable than probabilistic knowledge about the square root of the chance of heads (which you lack, and your friend has). But there *is* an important difference between you. Given that you both satisfy the Principal Principle, *your* credence that the coin will in fact come up heads on the next toss — call this proposition 'HEADS' — is also the product of cognitive ability, as we will see. Hers is less so. If this is right, then there is good epistemic reason to describe your inference problem in terms of B (*i.e.*, as a problem of adjudicating

[75]The MaxSen prior s over hypotheses of the form $B = x$ is equivalent to the non-symmetric prior over hypotheses of the form $\theta = x$, defined by the probability density $h(x) = 2.9469x(x^2 - x^4)^{0.2}$(figure 17).

between hypotheses about the coin's bias, or chance of HEADS), rather than $\theta = \sqrt[2]{B}$, or any other quantity. Describing your inference problem in this way does the most to help you secure probabilistic knowledge about whether the coin will *in fact* come up heads on the next toss.

It is not too hard to see why. Your credence for HEADS is tied to your credences about the coin's bias, or the chance of HEADS, if you are rational, in a way that it is not tied to your credences about the square root of its bias (or any other quantity). Minimally rational credence functions satisfy the Principal Principle. And given that your credence function satisfies the Principal Principal, you estimate the truth-value of HEADS by estimating its chance. Your credence (truth-value estimate) for HEADS just is your best estimate (expectation) of the chance of HEADS. Your credence for HEADS is not, in contrast, determined by your best estimate (expectation) of $\theta = \sqrt[2]{B}$, $\delta = cos\, B$, or any other function of the coin's bias B. It is a well-known fact that one's expected value for B is *not* a function of one's expected value for θ, or δ, etc. (see, *e.g.*, Pettigrew (2012)).[76] Best estimates (expectations) are simply not cleanly related in this way. So neither is your credence for HEADS a function of your best estimate (expectation) of θ, δ, etc.

This tells us something about how to *explain* the accuracy of your posterior credence for HEADS. Given that you satisfy the Principal Principal, and so estimate truth-values by estimating chances, two facts explain its accuracy: (i) how close your posterior chance estimate was to the true chance of heads (how *accurate* your posterior chance estimate was), and (ii) whether events went as expected; whether the truth-value of HEADS happened to fall close to (or far from) from its chance.

This in turn tells us something about how your *evidence* might explain the accuracy of your posterior credence for HEADS. The character of your evidence (about the 5 initial

[76]To see this, consider a concrete example. Suppose you adopt the MaxSen prior over hypotheses about the coin's bias, B, *viz.*, the beta distribution s with $\alpha = \beta \approx 1.2$. Then your expected value for B is 0.5, while the square of your expected value for θ is approximately 0.453.

flips) does not explain the latter fact, *viz.*, why the truth-value of HEADS happened, on this occasion, to fall close to (far from) from its chance. It has no bearing on the world's falling in line with chance or not. So it must explain the former fact, if it is to be relevant to the accuracy of your posterior credence at all. That is, it must explain the accuracy of your *posterior chance estimate* if it is to explain the accuracy of your posterior credence for HEADS. And it *must* do this last bit, if that credence is to count as skillfully produced, and hence be a candidate for constituting probabilistic knowledge.

At bottom, this is why describing your inference problem in terms of B — framing it as a problem that requires sorting out the true bias, or chance of HEADS — rather than $\theta = \sqrt[2]{B}$ (or any other quantity) helps you secure probabilistic knowledge about whether the coin will *in fact* come up heads on the next toss. For your posterior credence in HEADS to constitute probabilistic knowledge, your evidence must explain its accuracy to the greatest degree possible. And it must do so by explaining the accuracy of the underlying *chance estimate* that determines it. Now recall that our summary statistic — variance in objective expected posterior accuracy — tracks the extent to which one's evidence explains one's posterior accuracy (and, in turn, the extent to which one's posterior accuracy is the product of cognitive ability). The smaller that statistic, the greater the extent to which the character of the evidence explains that accuracy. So the objective expected posterior accuracy of the chance estimate that determines your credence in HEADS must vary minimally across chance hypotheses. Finally — and this is the crucial point — this happens exactly when you describe your inference problem in terms of the bias of the coin, B.

To see this, measure the inaccuracy of your posterior chance estimate — *i.e.*, the expected value of the coin's bias B determined by your posterior credence function c': $Exp_{c'}(B)$ — by one particularly attractive scoring rule, *viz.*, the Brier score (§5.2). And measure variance in objective expected posterior accuracy across chance hypotheses, $var(Exp_{c'}(B))$, in the usual way (by the difference in its maximum and minimum expected inaccuracies). Now note that $var(Exp_{c'}(B))$ takes a unique minimum when your prior credences over hypotheses about the coin's bias, B, are given by the beta distribution s with $\alpha = \beta \approx 1.2$ (figure 18). And MaxSen delivers s as your prior exactly when you describe your inference problem in terms of B, rather than $\theta = \sqrt[2]{B}$, or any other quantity.[77]

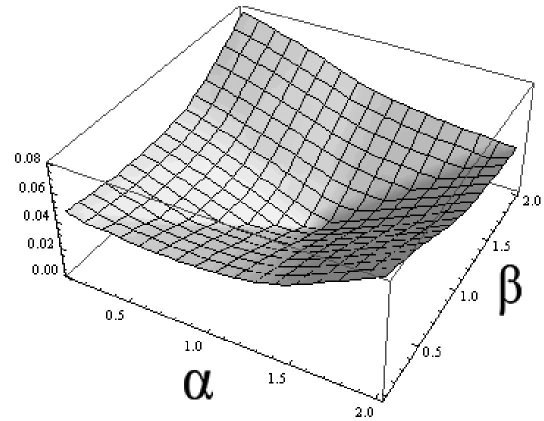

Figure 18: $var(Exp_{c'}(B))$ for beta priors c with $0 \leq \alpha, \beta \leq 2$, given n=5.

The moral: MaxSen is language or description dependent, like MaxEnt. But it also furnishes a clear rationale for favoring one way of describing your inference problem over others. The rationale: there is a single description that does the most to help secure probabilistic knowledge not simply about theoretical hypotheses (*e.g.*, hypotheses about the chances), but also about non-theoretical propositions (*e.g.*, whether the coin will in fact come up heads). So MaxSen is not inconsistent, in the way MaxEnt seems to be. It does not yield different prescriptions relative to different ways of describing your inference problem while staying silent about how to choose between them.

10 Appendix B: Outstanding Issues

This paper outlines one kind of formal tool — the maximum sensitivity principle, or MaxSen — for securing skillfully produced credences. But it does not provide a full explication (or defense) of MaxSen. For example, we restricted our attention to inference

[77] While $var(Exp_{c'}(B))$ takes a minimum at $c = s$ *in this particular inference problem*, it is an open question whether this is true more generally.

problems involving simple theoretical hypotheses about the bias of a coin and binomial data, *i.e.*, data that comes in the form of a sequence of 'successes' (heads) and 'failures' (tails), and is generated by a fixed number of identical and independent trials of an experiment (I.I.D. trials).

As it turns out, we can lift this restriction. MaxSen generalizes straightforwardly to inference problems involving more complicated theoretical hypotheses and data, *e.g.*, multinomial chance hypotheses — hypotheses about the chance of not only two possible outcomes (*e.g.*, heads or tails), but more generally, any finite number of possible outcomes — and multinomial data — sequences of such outcomes generated by a fixed number I.I.D trials. (Indeed, MaxSen generalizes straightforwardly to inference problems involving *any* parametric family of chance hypotheses, *i.e.*, hypotheses about chance distributions that can be described by a finite number of parameters.) And quite a few inference problems in the sciences, engineering, etc., are like this. They involve performing a fixed number of identical and independent trials of an experiment with only finitely many possible outcomes, in order to adjudicate between competing theoretical hypotheses, which specify the respective chances of those outcomes. Any problem of this sort can be thought of as sampling from a multinomial chance distribution (*cf.* Walley (1996)). Here, for example, are two paradigmatic instances:

- An aerospace engineering company tests a number of nearly identical rocket prototypes, and records which of the (finitely many) components fail in any unsuccessful run, in order to adjudicate between competing hypotheses about the chance of such a rocket failing in one way or another.

- A pharmacology lab runs a double-blind, randomized, controlled study of a new drug, and records any of the (finitely many) possible side-effects, in order to adjudicate between competing hypotheses about the drug's impact on one's chances of suffering such side-effects.

Of course, many inference problems are *not* like this. Microbiologists, for example, design and perform experiments aimed at adjudicating between even more complex theoretical hypotheses than those discussed above, *e.g.*, Hidden Markov models that describe the various networks (protein-protein interaction networks, etc.) that give rise to cell behavior (*cf.* Barabasi and Oltvai (2004)). In inference problems of this sort, prior credences will turn out to be enormously complex (nonparametric priors).[78] A fuller explication of MaxSen would illustrate how inference techniques familiar from nonparametric Bayesian statistics and machine learning (*e.g.*, Markov Chain Monte Carlo) can be used to determine a MaxSen prior in such problems, and to compute its posterior.[79]

In addition, in some contexts of inquiry, we simply do not know enough about the data-generating process to know which data items our experiment might produce (what statisticians call the *sample space* of the experiment). For example, while you might know enough about drug A to know what its possible side-effects are (because it is relevantly similar to other well-studied drugs), you might have no idea what the possible side-effects of drug B are. And in still other contexts of inquiry, new scientific data is simply not generated by running I.I.D. trials of an experiment, or anything of the sort, *e.g.*, in paleontology, geology, etc. In such contexts, it is not obvious how to generalize MaxSen to help secure probabilistic knowledge.

Examining the boundaries of the class of contexts in which MaxSen is applicable, and responding in full to language dependence concerns are tasks that require separate investigation. Our aim here was simply to illustrate, in broad strokes, one kind of formal tool for delivering skillfully produced credences, and in doing so, to provide a framework for theorizing about how to secure probabilistic knowledge. I conclude by raising a few

[78]Nonparametric priors specify a joint distribution over an infinite number of parameters, *e.g.*, each of the uncountably many values of a probability density function.

[79]See Orbanz and Teh (2010) for an overview of nonparametric Bayesian statistics.

additional questions to be addressed in future research.

- We specified the MaxSen prior using one particular proper scoring rule, or inaccuracy measure, *viz.*, the continuous ranked probability score. How robust are our results across other proper scoring rules, *e.g.*, the energy score (Gneiting and Raftery (2007, p. 367))?

- MaxSen minimizes the explanatory relevance of one particular non-evidential factor, *viz.*, prior accuracy. Are there any other sorts of non-evidential factors whose explanatory relevance we might mitigate by savvy prior construction? What sorts of hurdles might we face when handling multiple factors at once? What sorts of compromises might we be forced to strike?

- The MaxSen prior outperforms alternative precise priors, including the MaxEnt prior, vis-à-vis delivering skillfully produced posteriors. But we have not compared the MaxSen prior to *imprecise* priors, or sets of probability functions. Is there good epistemic reason to prefer the MaxSen prior to alternative imprecise priors, at least in certain contexts of inquiry?*

References

Barabasi, A.-L. and Z. N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* 5, 101–113.

Bennett, J. 1988. *Events and their Names*. Indianapolis: Hackett Publishers.

*Thanks to Joanna Burch-Brown, Chris Burr, Steve Campbell, Dmitri Gallow, Allan Gibbard, Martha Gibson, Pavel Janda, Jim Joyce, Ben Levinstein, Sarah Moss, Samir Okasha, Richard Pettigrew, Patrick Shirreff, Dan Singer, Dennis Stampe, Rohan Sud, Mike Titelbaum, audiences at UW Madison and the MCMP, as well as anonymous referees for *The Philosophical Review* for helpful comments and advice. I was supported by the ERC Starting Grant 'Epistemic Utility Theory: Foundations and Applications' during my work on this paper.

Birnbaum, A. 1962. On the foundations of statistical inference. *Journal of the American Statistical Association 57*(298), 269–306.

Easwaran, K. 2013. Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection. *Philosophy of Science 80*(1), 119–142.

Edwards, W., H. Lindman, and L. J. Savage. 1963. Bayesian Statistical Inference for Psychological Research. *Pyschological Review 70*(3), 193–242.

Eriksson, L. and A. Hajek. 2007. What are degrees of belief? *Studia Logica 86*(2), 183–213.

Fetzer, J. 1982. Probabilistic explanations. In *PSA: Proceedings of the Biennial Meeting of Philosophy of Science Association*, Volume 2, pp. 194–207.

Fetzer, J. 1983. Probability and objectivity in deterministic and indeterministic situations. *Synthese 57*, 367–386.

Fisher, R. 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 222*, 309–368.

Foley, R. 1992. The epistemology of belief and the epistemology of degrees of belief. *American Philosophical Quarterly 29*(2), 111–124.

Furrer, F., J. Aberg, and R. Renner. 2011. Min- and max-entropy in infinite dimensions. *Communications in Mathematical Physics 306*(1), 165–186.

Gillies, D. 2000. Varieties of propensity. *British Journal for the Philosophy of Science 51*, 807–835.

Gneiting, T. and A. Raftery. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Greaves, H. and D. Wallace. 2006. Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind 115*(459), 607–632.

Greco, J. and J. Turri. 2013. Virtue epistemology. *The Stanford Encyclopedia of Philosophy http://plato.stanford.edu/archives/win2013/entries/epistemology-virtue/.*

Hersbach, H. 2000. Decomposition of the continuous ranked probabil- ity score for ensemble prediction systems,. *Weather and Forecasting 15*, 559–570.

Hitchcock, C. 2012. Cause and chance.

Jaynes, E. T. 1957. Information theory and statistical mechanics. *Physical review 106*(4), 620.

Jeffrey, R. 1965. *The Logic of Decision*. University of Chicago Press.

Jeffrey, R. 1970. Dracula meets wolfman: Acceptrance vs. partial belief. In M. Swain (Ed.), *Induction, Acceptance, and Rational Belief*, pp. 157–185. Dordrecht: Reidel.

Jeffrey, R. 2004. *Subjective Probability: The Real Thing*. Cambridge: Cambridge University Press.

Joyce, J. 1998. A nonpragmatic vindication of probabilism. *Philosophy of Science 65*, 575–603.

Joyce, J. 2005. How probabilities reflect evidence. *Philosophical Perspectives 19*, 153–178.

Joyce, J. 2007. VIII—Epistemic Deference: The Case of Chance. *Proceedings of the Aristotelian Society CVII*(Part 2), 187–206.

Joyce, J. 2009. Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In F. Huber and C. Schmidt-Petri (Eds.), *Degrees of Belief*, Volume 342. Dordrecht: Springer.

Joyce, J. 2010. A defense of imprecise credences in inference and decision making. *Philosophical Perspectives 24*, 281–323.

Kaplan, M. 2010. In defense of modest probabilism. *Synthese 176*(1), 41–55.

Keynes, J. M. 1921. *A Treatise on Probability*. London: Macmillan and Co.

Kraft, C. H., J. W. Pratt, and A. Seidenberg. 1959. Intuitive probability on finite sets. *The Annals of Mathematical Statistics 30*(2), 408–419.

Krantz, D., D. Luce, P. Suppes, and A. Tversky. 1971. *Foundations of Measurement: Volume 1, Additive and Polynomial Representations*. New York and London: Academic Press, Inc.

Kyburg, H. E. 1961. *Probability and the Logic of Rational Belief*. Connecticut: Wesleyan University Press.

Leitgeb, H. 2013. The stability theory of belief. *Philosophical Review 123*(2), 131–171.

Leitgeb, H. and R. Pettigrew. 2010. An objective justification of Bayesianism II: The consequences of minimizing inaccuracy*. *Philosophy of Science 77*(2), 236–272.

Levi, I. 1980. *The Enterprise of Knowledge*. Cambridge, MA: MIT Press.

Lewis, D. 2000. Causation as Influence. *The Journal of Philosophy 97*(4), 182–197.

List, C. and M. Pivato. 2015. Emergent chance. *Philosophical Review 124*(1), 119–152.

Meacham, C. J. G. 2005. Three Proposals regarding a Theory of Chance. *Noûs 39*, 281–307.

Meacham, C. J. G. 2010. Two Mistakes Regarding the Principal Principle. *The British Journal for the Philosophy of Science 61*(2), 407–431.

Mellor, D. H. 1995. *The Facts of Causation*. London: Routledge Press.

Moss, S. 2013. Epistemology Formalized. *Philosophical Review 122*(1), 1–43.

Nozick, R. 1981. *Philosophical Explanations*. Oxford: Oxford University Press.

Oddie, G. 1997. Conditionalization, cogency, and cognitive value. *The British Journal for the Philosophy of Science 48*(4), 533–541.

Orbanz, P. and Y. W. Teh. 2010. Bayesian nonparametric models. In *Encyclopedia of Machine Learning*, pp. 81–89. Springer.

Pettigrew, R. 2012. Accuracy domination arguments and credences as estimates of truth-values. *Manuscript*.

Pettigrew, R. 2013a. Epistemic utility and norms for credences. *Philosophy Compass 8*(10), 897–908.

Pettigrew, R. 2013b. A new epistemic utility argument for the Principal Principle. *Episteme 10*(01), 19–35.

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

Predd, J. B., R. Seiringer, E. H. Lieb, D. N. Osherson, H. V. Poor, and S. R. Kulkarni. 2009. Probabilistic Coherence and Proper Scoring Rules. *IEEE Transactions on Information Theory 55*(10), 4786–4792.

Pritchard, D. 2005. *Epistemic Luck*. Oxford: Clarendon.

Pritchard, D. 2008. Sensitivity, safety, and anti-luck epistemology. In J. Greco (Ed.), *The Oxford Handbook of Scepticism*, pp. 437–55. Oxford: Oxford University Press.

Pritchard, D. 2010. Anti-luck virtue epistemology. In D. P. A. Haddock, A. Millar (Ed.), *The Nature and Value of Knowledge: Three Investigations*. Oxford: Oxford University Press.

Pritchard, D. 2012. Anti-luck virtue epistemology. *Journal of Philosophy 109*(3), 247.

Savage, L. 1954. *The Foundations of Statistics*. New York: Dover.

Schervish, M., T. Seidenfeld, and J. Kadane. 2009. Proper scoring rules, dominated forecasts, and coherence. *Decision Analysis 6*(4), 202–221.

Scott, D. 1964. Measurement structures and linear inequalities. *Journal of mathematical psychology 1*(2), 233–247.

Seidenfeld, T. 1986. Entropy and uncertainty. *Philosophy of Science 53*(4), 467–491.

Singer, P. 1975. *Animal Liberation*. New York: A New York Review Book.

Singer, P. 1985. Ethics and the new animal liberation movement. In P. Singer (Ed.), *In Defense of Animals*, pp. 1–10. New York: Basil Blackwell.

Sosa, E. 1999. How must knowledge be modally related to what is known? *Philosophical Topics* (26), 373–384.

Sosa, E. 2007. *Apt Belief and Reflective Knowledge, Volume 1: A Virtue Epistemology*. Oxford: Oxford University Press.

Stocker et al. 2013. IPCC, 2013: Summary for policymakers. In Stocker et al. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group 1 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 1–30. Cambridge: Cambridge University Press.

Swanson, E. 2006. Interactions With Context. *PhD Diss.*, 1–119.

Titelbaum, M. 2015. *Fundamentals of Bayesian Epistemology*. Oxford: Oxford University Press.

Venn, J. 1866. *The Logic of Chance*. London: Macmillan.

Walley, P. 1996. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society. Series B (Methodological) 58*(1), 3–57.

Williamson, J. 2010. *In Defence of Objective Bayesianism*. Oxford University Press.

Zagzebski, L. 1996. *Virtues of the Mind*. Cambridge: Cambridge University Press.